(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/097060 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/18354

(22) International Filing Date: 22 May 2002 (22.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/293,768	25 May 2001 (25.05.2001)	US
60/309,548	1 August 2001 (01.08.2001)	US
60/314,400	23 August 2001 (23.08.2001)	US
60/343,706	19 October 2001 (19.10.2001)	US
60/337,999	7 December 2001 (07.12.2001)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **GORVAD, Ann, E.** [US/US]; 369 Marie Common, Livermore, CA 94550 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue # 71, Palo Alto, CA 94303 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **YUE, Huibin** [US/US]; 1170 South Stelling Road, Cupertino, CA 95014 (US). **FORSYTHE, Ian, J.** [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). **LI, Joana, X.** [US/US]; 1264 Geneva Avenue, San Francisco, CA 94112 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street A # 205, San Leandro, CA 94577 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **MASON, Patricia, M.** [US/US]; 360 Clarke Lane, Morgan Hill, CA 95014 (US). **LAL, Preeti, G.** [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **LEE, Sally** [US/US]; 3643 26th Street, San Francisco, CA 94110 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). **TANG, Y. Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/097060 A2

(54) Title: CARBOHYDRATE-ASSOCIATED PROTEINS

(57) Abstract: The invention provides human carbohydrate-associated proteins (CHOP) and polynucleotides which identify and encode CHOP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of CHOP.

CARBOHYDRATE-ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of carbohydrate-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of carbohydrate metabolism, cell proliferative, autoimmune/inflammatory, reproductive, genetic, transport, and neurological disorders and cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of carbohydrate-associated proteins.

BACKGROUND OF THE INVENTION

Carbohydrates, including sugars or saccharides, starch, and cellulose, are aldehyde or ketone compounds with multiple hydroxyl groups. Carbohydrates have three important roles in mammalian cells. Carbohydrates function as energy-storage molecules, as fuels, and as metabolic intermediates. Carbohydrates are broken down to release energy in glycolysis or may be stored as glycogen for later use. The importance of carbohydrate metabolism is demonstrated by the sensitive regulatory system in place for maintenance of blood glucose levels. Two pancreatic hormones, insulin and glucagon, promote increased glucose uptake and storage by cells, and increased glucose release from cells, respectively. The sugars deoxyribose and ribose form part of the structural support of DNA and RNA, respectively, providing a second example of carbohydrate function. Third, carbohydrates provide a means for post-translational modification of secreted and membrane proteins and lipids. Indeed, 2-10% of the content of eukaryotic cell membranes are contributed by oligosaccharides on membrane glycoproteins and glycolipids. Carbohydrate modifications on glycoproteins and glycolipids create great structural diversity, and since they are mainly located on the extracellular side of the plasma membrane, they play an important role in intercellular recognition (Stryer, L. (1988) Biochemistry, W.H. Freeman and Company, New York NY, pp. 298-299, 331-347).

Proteins are associated with carbohydrates in several ways. Carbohydrate-containing macromolecules, which include glycoproteins, glycolipids, glycosaminoglycans, and proteoglycans, are found on the cell surface and in the extracellular matrix. The extracellular matrix is composed of diverse glycoproteins, and carbohydrate-binding proteins which are secreted from the cell and assembled into an organized meshwork in close association with the cell surface. The interaction of the cell with the surrounding matrix profoundly influences cell shape, strength, flexibility, motility, and adhesion. These dynamic properties are intimately associated with signal transduction pathways controlling cell proliferation and differentiation, tissue construction, and embryonic development.

Glycoproteins have covalently attached carbohydrates which have been added to the proteins as they traverse the secretory pathway. Some proteins noncovalently associate with carbohydrate-

containing macromolecules for purposes of binding, modifying, or degrading the carbohydrates. Glycoproteins include cell adhesion molecules, receptors, blood group antigens, growth factors, and antibodies. These proteins are involved in cellular processes such as cell-cell recognition and signaling, recognition and/or destruction of neurotransmitters, transmission of neural impulses, and immune function.

Oligosaccharide modifications can provide great structural diversity. N- and O-linked oligosaccharides are transferred to proteins and modified in a series of enzymatic reactions that occur in the endoplasmic reticulum (ER) and Golgi. Oligosaccharides stabilize the protein during and after folding, orient the protein in the membrane, improve the protein's solubility, and act as a signal for lysosome targeting.

Heavily glycosylated glycoproteins are also referred to as proteoglycans. Proteoglycans in the extracellular matrix of connective tissues such as cartilage are essential for distributing the load in weight-bearing joints. Cell-surface-attached proteoglycans anchor cells to the extracellular matrix. Both extracellular and cell-surface proteoglycans bind growth factors, facilitating their binding to cell-surface receptors and subsequent triggering of signal transduction pathways (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York NY, pp. 1139-1142).

Carbohydrates also form glycosaminoglycans (GAGs), which are linear unbranched polysaccharides composed of repetitive disaccharide units. GAGs exist free or as part of proteoglycans, large molecules composed of a core protein attached to one or more GAGs. GAGs are found on the cell surface, inside cells, and in the extracellular matrix. The GAG hyaluronan (HA) is found in the extracellular matrix of many cells, especially in soft connective tissues, and is abundant in synovial fluid (Pitsillides, A.A. et al. (1993) Int. J. Exp. Pathol. 74:27-34). HA, which functions in water and plasma protein homeostasis, seems to play important roles in cell regulation, development, and differentiation.

Glycolipids, along with phospholipids and cholesterol, form the membranes of cells. Examples of glycolipids include blood group antigens on erythrocytes and gangliosides in the myelin sheath of neurons. Modifications to glycoproteins and glycolipids on the extracellular side of the plasma membrane are important for intercellular recognition (Stryer, *supra*, pp. 298-299, 331-347; Lodish, et al., *supra*, pp. 612-615).

Lectins are extracellular glycoproteins which bind cell surface carbohydrates specifically and reversibly, resulting in the agglutination of cells (Drickamer, K. and Taylor, M.E. (1993) Annu. Rev. Cell Biol. 9:237-264). This function is particularly important for activation of the immune response. Lectins mediate the agglutination and mitogenic stimulation of lymphocytes at sites of inflammation (Lasky, L.A. (1991) J. Cell. Biochem. 45:139-146; Paietta, E. et al. (1989) J. Immunol. 143:2850-2857).

Lectins are classified into subfamilies based on carbohydrate-binding specificity. The galectin subfamily, in particular, includes lectins that bind β-galactoside carbohydrate moieties in a thiol-dependent manner (Hadari, Y.R. et al. (1995) J. Biol. Chem. 270:3447-3453). Galectins are widely expressed and developmentally regulated. Because all galectins lack an N-terminal signal peptide, it is suggested that galectins are externalized through an atypical secretory mechanism. Two classes of galectins have been defined based on molecular weight and oligomerization properties. Galectins contain a characteristic carbohydrate recognition domain (CRD), also known as a galaptin domain, which is about 140 amino acids long and contains several conserved residues (See Prosite PDOC00279 Vertebrate galactoside-binding lectin signature).

Another example is intelectin, a Ca^{2+} dependent lectin that binds to galactofuranosyl residues and bacterial arabinogalactan. Intelectin may play a role in the recognition of bacterial carbohydrate and induction of the immune response to microorganisms.

Carbohydrate-modifying enzymes

The enzyme glutamine:fructose-6-phosphate amidotransferase (GFAT), also known as aminotransferase, catalyzes the reversible reaction of L-glutamine and D-fructose-6-phosphate to form L-glutamate and D-glucosamine-6-phosphate, which is the rate-limiting step in the hexosamine biosynthetic pathway (ExPASy ENZYME: EC 2.6.1.16). D-glucosamine-6-phosphate acts in the biosynthesis of UDP-N-acetyl-glucosamine (UDP-GlcNAc) and other hexosamines incorporated into glycoproteins and proteoglycans. GFAT regulates the availability of precursors for N- and O-linked glycosylation. Glucosamine enhances the production of transforming growth factor (TGF)-β1 (Kolm-Litty, V. et al. (1998) J. Clin. Invest. 101:160-169). GFAT activity plays a role in insulin resistance in Type II diabetes, and GFAT overexpression leads to insulin resistance. Hexosamine metabolism appears to regulate glycogen synthase, the rate-limiting enzyme in glycogen synthesis, as well as PP1G, a glycogen-bound protein phosphatase, pyruvate kinase, and the glucose transporter GLUT1 (McClain, D.A. and Crook, E.D. (1996) Diabetes 45:1003-1009).

The enzyme glucosamine-6-phosphate deaminase (GNPDA), also known as isomerase, catalyzes the reversible reaction of D-glucosamine-6-phosphate with water to form D-fructose-6-phosphate and ammonia (ExPASy ENZYME EC 5.3.1.10). This reaction links hexosamine systems with glycolytic pathways and may provide an energy source from the catabolism of hexosamines in glycoproteins, glycolipids, and sialic-acid-containing macromolecules. GNPDA is expressed in tissues with high energy requirements (Wolosker, H. et al. (1998) FASEB J. 12:91-99).

The enzyme UDP-glucose dehydrogenase (UDPGD) catalyzes the reversible reaction of UDP-glucose, 2 NAD^+, and water to form UDP-glucuronate and 2 NADH (ExPASy ENZYME EC 1.1.1.22). UDP-glucuronate is needed for the biosynthesis of GAGs, which appear to play a role in signal transduction pathways (Binari, R.C. et al. (1997) Development 124:2623-2632).

Man_9-mannosidase is an α1,2-mannosidase (glycosyl hydrolase) involved in the early processing of N-linked oligosaccharides. This enzyme catalyzes the specific cleavage of α1,2-mannosidic linkages in Man_9-$(GlcNAc)_2$ and Man_5-$(GlcNAc)_2$. Multiple α1,2-mannosidases have been identified in mammalian cells and may be needed for the processing of distinct classes of N-glycoproteins. Man_9-mannosidase is a Type II membrane protein with a short cytoplasmic tail, a single transmembrane domain, and a large luminal catalytic domain. The human kidney enzyme is localized to the Golgi (Bause, E. et al. (1993) Eur. J. Biochem. 217:535-540; Bieberich, E. and Bause, E. (1995) Eur. J. Biochem. 233:644-649).

DPM1 is an enzyme in the endoplasmic reticulum that catalyzes the production of dolichol phosphate-mannose (DPM) from GDP-mannose and dolichol phosphate. The activity of DPM1 is regulated by DPM2, which targets DPM1 to the endoplasmic reticulum (ER) and increases its affinity for dolichol phosphate. DMP2 resides in the (ER) membrane and contains two putative transmembrane domains and a putative ER-localization signal near its C-terminus.

Glycosylation

Glycosylation refers to the covalent attachment of any number of carbohydrate chains (oligosaccharides) to specific sites (glycosylation sites) on proteins. Glycosylation is a post-translational protein modification essential to the conformation, stability, transport, secretion, antigenicity, clearance and activity of glycosylated proteins (glycoproteins). The composition of the attached oligosaccharides is specific to a protein and may be simple (consisting primarily of mannose residues) or complex (with additional N-acetyl-glucosamine (GlcNAc), sialic acid, and galactose residues). Glycoproteins may have relatively few carbohydrate groups or may contain a larger percentage of carbohydrate than protein (based on molecular weight). These latter, heavily glycosylated glycoproteins are also referred to as proteoglycans to emphasize the predominant carbohydrate composition of the molecules. The type of saccharide bond (e.g., α, β, 1,2-, 1,4-) formed between any two constituent carbohydrate residues is also a critical molecular determinant for the structure and function of the glycoprotein.

Glycosylation confers increased hydrophilicity to proteins. Many glycoproteins, such as carrier proteins, antibodies, and lysosomal proteins, are found free in solution (e.g., plasma). Other glycoproteins are membrane-bound. In the case of membrane-associated glycoproteins, the carbohydrate side-chain serves to orient the glycoproteins in the membrane lipid bilayer. The glycosylated regions of the molecule interact with the aqueous environment on the inside or outside of the membrane while the more hydrophobic domains of the glycoprotein (typically consisting of non-polar amino acid residues that are not glycosylated) interact with the phospholipids in the membrane.

Addition of oligosaccharide side chains occurs at the $-NH_2$ group of asparagine (Asn)

residues (N-linked glycosylation) or at the -OH group of serine (Ser) residues (O-linked glycosylation). The process of N-linked glycosylation begins in the endoplasmic reticulum (ER) and is completed in the Golgi apparatus of eukaryotic cells. O-linked glycosylation occurs exclusively in the Golgi. Of all characterized glycoproteins, 90% are N-glycosylated, with or without additional O-glycosylation. Only 10% are exclusively O-glycosylated (Apweiler, R. et al. (1999) Biochim. Biophys. Acta 1473:4-8). Almost two-thirds of the approximately 75,000 SWISS-PROT protein sequences include putative N-glycosylation sites, underscoring the importance of this protein modification in nature (Apweiler et al., *supra*). Biochemical steps involved in N-linked glycosylation have been well characterized, and are reviewed below.

N-linked glycosylation

N-linkage of carbohydrates to proteins occurs via a nitrogen atom of asparagine (Asn) residue side-chains in the context of the primary amino acid sequence Asn-X-Ser or Asn-X-Thr (Ser = serine, Thr = threonine, and X = any amino acid residue except proline). While the composition of N-linked oligosaccharides is highly diverse, the pathways responsible for glycosylation have common first steps. A 14-residue core oligosaccharide, containing two N-acetylglucosamine (GlcNAc), nine mannose, and three glucose residues, is transferred as a unit from a dolichol phosphate donor molecule to the $-NH_2$ group of an acceptor Asn residue on the target protein. Typically, the three glucose residues of the core oligosaccharide are removed by glucosidases I and II resulting in "high mannose oligosaccharides" side chains. These partially processed N-linked glycoproteins are then sequentially transported from the ER through the *cis*-, *medial*-, and *trans*-cisternae of the Golgi (Bonay, P. et al. (1996) J. Biol. Chem. 271:3719-3726). Further modification to the oligosaccharide chains may occur to remove additional core mannose residues using the enzymes Golgi mannosidase I (*cis*-cisterna), N-acetyl-glucosaminyltransferase (GluNAcT; *medial*-cisterna), and Golgi mannosidase II (*trans*-cisternae). Following the removal of some of the mannose residues by Golgi mannosidase I, the addition of a single GlcNAc by GluNAcT is essential for the removal of the remaining mannose residues of the core oligosaccharide by Golgi mannosidase II.

Mannose-1-phosphate guanyltransferases are involved in early steps of protein glycosylation. They participate in sugar metabolism and their enzymatic products are channeled into glycoprotein synthesis. Mannose-1-phosphate guanyltransferase 1 (MPG1), also referred to as NDP-hexose pyrophosphorylase, catalyzes the conversion of GTP and αD-mannose 1-phosphate into diphosphate and CDP-ethanolamine in mannose metabolism. This enzyme is very similar to CDP-glucose pyrophosphorylase and may also be involved in the regulation of cell cycle progression. A cDNA coding for GTP:αD-mannose-1-phosphate guanyltransferase 1 (MPG1) was recently isolated from a cDNA library of a *Trichoderma reesei* strain (Kruszewska, J.S. et al. (1998) Curr. Genet. 33:445-50). The nucleotide sequence of the 1.6 kb cDNA revealed an ORF which encodes a protein of 364 amino

acids. Sequence comparisons demonstrate 70% identity with the yeast *Saccharomyces cerevisiae* guanyltransferase gene 1 (MPG1) and 75% identity with the *Schizosaccharomyces pombe* homologue.

Complex oligosaccharide side-chains result from the addition of N-acetyl-glucosamine, N-acetylneuraminic acid (sialic acid), and galactose, as well as other sugar moieties, to the remaining core sugar moieties on the partially-processed glycoprotein. These modifications occur in the *trans*-cisterna and *trans*-Golgi network (TGN), and involve a number of enzymes including N-acetyl-glucosaminyltransferase I (GlcNAcTsI), sialyltransferases (STs), and galactosyltransferases (GalTs). Multiple isoforms of many of these enzymes produce specific α or β, 1,2-, 1,3-, 1,4-, or 1,6-disaccharide bonds between constituent sugar residues of the oligosaccharide side-chain. The stereochemistry and type of bonds in a carbohydrate side-chain contribute to the overall structure and function of the resulting glycoprotein (Lehninger, A. L. et al. (1993) Principles of Biochemistry, Worth Publishers, New York NY, pp. 931; Lewin, B. (1997) GenesVI, Oxford University Press, New York NY, pp. 1030-1033).

Galactosyltransferases are a subset of glycosyltransferases that transfer galactose (Gal) to the terminal N-acetylglucosamine (GlcNAc) oligosaccharide chains that are part of glycoproteins or glycolipids that are free in solution (Kolbinger, F. et al. (1998) J. Biol. Chem. 273:433-440; Amado, M. et al. (1999) Biochim. Biophys. Acta 1473:35-53). Galactosyltransferases are found in the Golgi, on the cell surface, and as soluble extracellular proteins, in addition to being present in the Golgi. β1,3-galactosyltransferases form Type I carbohydrate chains with Gal (β1-3)GlcNAc linkages. β1,3-galactosyltransferases appear to have a short cytosolic domain, a single transmembrane domain, and a catalytic domain with eight conserved regions (Kolbinger et al., *supra*; Hennet, T. et al. (1998) J. Biol. Chem. 273:58-65). In mouse, UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 1 is located at amino acid residues 78-83, region 2 is located at amino acid residues 93-102, region 3 is located at amino acid residues 116-119, region 4 is located at amino acid residues 147-158, region 5 is located at amino acid residues 172-183, region 6 is located at amino acid residues 203-206, region 7 is located at amino acid residues 236-246, and region 8 is located at amino acid residues 264-275. A variant of a sequence found within mouse UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 8 is also found in bacterial galactosyltransferases, suggesting that this sequence defines a galactosyltransferase sequence motif (Hennet et al., *supra*). Recent work suggests that brainiac protein is a β1,3-galactosyltransferase. (Yuan, Y. et al. (1997) Cell 88:9-11; Hennet et al., *supra*).

UDP-Gal:GlcNAc-1,4-galactosyltransferase (-1,4-GalT) catalyzes the formation of Type II carbohydrate chains with Gal (β1-4)GlcNAc linkages (Sato, T. et al. (1997) EMBO J. 16:1850-1857). A soluble form of the enzyme is formed by cleavage of the membrane-bound form. Amino acids

conserved among β1,4-galactosyltransferases include two cysteines linked through a disulfide-bond and a putative UDP-galactose binding site in the catalytic domain (Yadav, S.P. and Brew, K. (1990) J. Biol. Chem. 265:14163-14169; Yadav, S.P. and Brew, K. (1991) J. Biol. Chem. 266:698-703; Shaper, N.L. et al. (1997) J. Biol. Chem. 272:31389-31399). β1,4-galactosyltransferases have several specialized roles in addition to synthesizing carbohydrate chains on glycoproteins or glycolipids. A β1,4-galactosyltransferase functions as part of a heterodimer with α-lactalbumin in mammary lactose production. A β1,4-galactosyltransferase on the surface of sperm functions as a receptor that specifically recognizes the egg. Cell surface β1,4-galactosyltransferases also function in cell adhesion, cell recognition, cell/basal lamina interaction, and normal and metastatic cell migration (Shur, B. (1993) Curr. Opin. Cell Biol. 5:854-863; Shaper, J. (1995) Adv. Exp. Med. Biol. 376:95-104; Masri, K.A. et al. (1988) Biochem. Biophys. Res. Commun. 157:657-663).

Synthetases are another class of carbohydrate-modifying enzymes that have critical roles in proper cell funtioning. For example, production of sialylated glycoconjugates requires the synthesis of cytidine 5'-monophosphate N-acetylneuraminic acid (CMP-Neu5Ac or CMP-sialic acid), a reaction catalyzed by CMP-Neu5Ac synthetase (Munster, A.K. et al. (1998) Proc. Natl. Acad. Sci. USA 95:9140-9145). Sialic acids of cell surface glycoproteins and glycolipids contribute to proper structure and function in a variety of tissues. Sialyltransferases (STs) are a subset of glycosyltransferases that catalyze the transfer of sialic acid (from CMP-sialic acids) to the carbohydrate groups of glycoproteins and glycolipids. A variety of these Type II membrane proteins are present in the Golgi. Cloned members of this gene family share an N-terminal cytoplasmic tail region, a transmembrane region, and a large luminal region containing three sialyl motifs designated large (L), small (S), and very small (VS). The L-sialyl motif contributes to donor substrate binding and consists of eight invariant residues within a highly conserved stretch of 48-49 amino acids. The 23-amino acid S-sialyl motif contributes to the binding of both donor and acceptor substrates (Datta, A. et al. (1997) Indian J. Biochem. Biophys. 34:157-65). In the case of a representative sialytransferase ST3GalI (~350 amino acids in length), the L, S, and VS, regions correspond to amino acids 138-182, 264-286, and 309-321, respectively. Other cloned members of the family include ST6GalNAcI and ST8SiaI. ST6GalNAcI is larger than the other known sialyltransferases, and is composed of more than 500 amino acid residues (Tsuji, S. et al. (1996) Glycobiology (letter) 6:v-vii; Geremia, R. et al. (1997) Glycobiology (letter) 7:v-vii; Datta, A. et al. (1995) J. Biol. Chem. 270:1497-1500; Datta, A. et al. (1998) J. Biol. Chem. 273:9608-9618; Tsuji, S. et al. (1998) J. Biochem. 120:1-13). Sialyltransferases are not abundant in cellular extracts, but several have been cloned and expressed. At least one inhibitor has been synthesized (Horenstein, B. et al. (1996) J. Am. Chem. Soc. 118:10371-10379).

A variety of other enzymes that are involved in sugar metabolism participate directly or

indirectly in glycosylation, upstream of events that occur in the ER and Golgi. Many of these enzymes were originally identified in bacteria and plants and are less well characterized in humans; however, human homologues may exist that perform similar functions. For example, ADP-glucose pyrophosphorylases catalyze a very important step in the biosynthesis of α1,4-glucans (glycogen or starch) in bacteria and plants, namely the synthesis of the activated glucosyl donor, ADP-glucose, from glucose-1-phosphate and ATP. ADP-glucose pyrophosphorylases are tetrameric, allosterically-regulated enzymes. There are a number of conserved regions in the sequence of bacterial and plant ADP-glucose pyrophosphorylase subunits. Additionally, there are three regions which are considered signature patterns. The first two regions are N-terminal and have been proposed to be part of the allosteric and substrate-binding sites in the *Escherichia coli* enzyme. The third pattern corresponds to a conserved region in the central part of the enzymes.

Carbohydrate metabolism disorders

Carbohydrate metabolism is altered in several disorders. Diabetes mellitus is characterized by abnormally high blood glucose (hyperglycemia). Type I diabetes results from an autoimmune-related loss of pancreatic insulin-secreting cells. Type II diabetes results from insulin resistance and impaired insulin secretory response to glucose, and is associated with obesity. Hypoglycemia, or abnormally low blood glucose levels, has several causes including drug use, genetic deficiencies in carbohydrate metabolism enzymes, cancer, liver disease, and renal disease (Berkow, R. et al. (1992) The Merck Manual of Diagnosis and Therapy, Internet Edition, Section 8, Chapter 91, Diabetes Mellitus, Hypoglycemia).

Mutations in enzymes involved in protein glycosylation cause severe diseases. For example, alpha mannosidase mutations cause congenital dyserythropoietic anemia Type I and alpha B lysosomal mannosidosis (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc. New York NY, pp. 2092-2093; and Online Mendelian Inheritance In Man, 224100).

Glucosidases represent another class of carbohydrate-modifying enzymes that catalyze the release of glucose from carbohydrates through hydrolysis of the glycosidic link in various glucosides. The inherited disorder type I Gaucher disease, which is characterized by hematologic abnormalities, can be detected in a heterozygous or homozygous individual through an assay of leukocyte beta-glucosidase levels (Raghavan, S.S. et al. (1980) Am. J. Hum. Genet. 32:158-173). Patients with all three types of Gaucher disease exhibit a deficiency of an enzyme called glucocerebrosidase that catalyzes the first step in the biodegradation of glucocerebroside. In the brain, glucocerebroside arises from the turnover of complex lipids during brain development and the formation of the myelin sheath of nerves. In other tissues, glucocerebroside arises mainly from the biodegradation of old red and white blood cells.

Galectins play a number of roles in diseases and conditions associated with cell-cell and cell-matrix interactions. For example, certain galectins associate with sites of inflammation and bind to cell surface immunoglobulin E molecules. In addition, galectins may play an important role in cancer metastasis. Galectin overexpression is correlated with the metastatic potential of cancers in humans and mice. Moreover, anti-galectin antibodies inhibit processes associated with cell transformation, such as cell aggregation and anchorage-independent growth.

Galectin-8, also known as prostate carcinoma tumor antigen 1 (PCTA-1), is a novel galectin implicated in cancer progression (Su, Z.-Z. et al. (1996) Proc. Natl. Acad. Sci. USA 93:7252-7257). Galectin-8 is expressed in invasive prostate carcinomas and early-stage prostate cancers, but not in normal prostate or benign prostatic hypertrophic tissue.

Defects in carbohydrate metabolism are also associated with cancer. Reduced GAG and proteoglycan expression is associated with human lung carcinomas (Nackaerts, K. et al. (1997) Int. J. Cancer 74:335-345). The carbohydrate determinants sialyl-LewisA and sialyl-LewisX are frequently expressed on human cancer cells. These determinants, ligands for the cell adhesion molecule E-selectin, are involved in the adhesion of cancer cells to vascular endothelium and contribute to hematogenous metastasis of cancer (Kannagi, R. (1997) Glycoconj. J. 14:577-584). Alterations of the N-linked carbohydrate core structure of cell surface glycoproteins are linked to colon and pancreatic cancers (Schwarz, R.E. et al. (1996) Cancer Lett. 107:285-291). Reduced expression of the Sda blood group carbohydrate structure in cell surface glycolipids and glycoproteins is observed in gastrointestinal cancer (Dohi, T. et al. (1996) Int. J. Cancer 67:626-631).

Changes in glycosaminoglycan levels are associated with several autoimmune diseases. Both increases and decreases in various GAGs occur in patients with autoimmune thyroid disease and autoimmune diabetes mellitus. Antibodies to GAGs were found in patients with systemic lupus erythematosus and autoimmune thyroid disease (Hansen, C. et al. (1996) Clin. Exp. Rheum. 14:S59-S67). The glycosaminoglycan hyaluronan (HA) induces tumor cell adhesion and migration, and its small fragments are angiogenic. Serum HA is diagnostic of liver disease and various inflammatory conditions, such as rheumatoid arthritis. Interstitial edema caused by accumulation of HA may cause dysfunction in various organs (Laurent, T.C. and Fraser, J.R. (1992) FASEB J. 6:2397-2404). Hyaluronidase is an enzyme that degrades HA to oligosaccharides by catalyzing the random hydrolysis of 1,4-linkages between N-acetyl-β-D-glucosamine and D-glucuronate residues. Hyaluronidases may function in cell adhesion, infection, angiogenesis, and signal transduction. Hyaluronidases are associated with reproduction, cancer, and inflammation. Hyaluronidase activity is significantly elevated in prostate tumor tissue compared to that in both normal prostate and benign prostate hyperplasia (Lokeshwar, V.B. et al. (1996) Cancer Res. 56:651-657).

PH-20, a protein expressed in the mammalian testis and present on the plasma membrane of

mouse and human sperm, has hyaluronidase activity (Lin, Y. et al. (1994) J. Cell Biol. 125:1157-63). PH-20 enables sperm to penetrate the mammalian egg's outer layer, which consists of about 3,000 cumulus cells embedded in an extracellular matrix rich in HA. Penetration of this layer is an essential step in the fertilization process. PH-20 is also expressed in some tumor cells. Non-testicular mammalian hyaluronidases include the HYAL1 hyaluronidase, expressed in human serum, and lysosomal hyaluronidase HYAL2, expressed in many cells (Lepperdinger, G. et al. (1998) J. Biol. Chem. 273:22466-22470). HYAL2 may have a role in producing distinct HA fragments that can induce angiogenesis and the expression of enzymes involved in signal transduction pathways, such as nitric oxide synthase. A lysosomal-type hyaluronidase may degrade HA in lung fibroblasts in a cytokine-regulated process (Sampson, P.M. et al. (1992) J. Clin. Invest. 90:1492-1503). The venom of numerous animals including various snakes, bees, hornets, stone fish, platypus, scorpions, and lizards contain hyaluronidase. Venom hyaluronidase is thought to act as an aid in the diffusion of toxins.

A number of human diseases are linked to genetic or acquired deficiencies in protein glycosylation. Carbohydrate-deficient glycoprotein syndromes (CDGSs) include a host of alterations in glycosylation in a number of disorders and diseases. CDGSs are a group of hereditary multisystem disorders (Matthijs, G. et al. (1997) Nat. Genet. 16:88-92) causing severe psychomotor and mental retardation, as well as blood coagulation abnormalities seen in thrombosis, bleeding, or stroke-like episodes. The characteristic biochemical abnormality of CDGSs is the hypoglycosylation (N-linked) of glycoproteins (Freeze, H. and Aebi, M. (1999) Biochim. Biophys. Acta 1455:167-78). Depending on the type of CDGS, the carbohydrate side chains of glycoproteins are either truncated or completely missing from the protein core. Several different types of CDGS have been classified. The most common form, CDGS type 1A, is caused by phosphomannomutase (PMM1) deficiency (Matthijs, G. (1998) Am. J. Hum. Genet. 62:542-50). PMM1 functions upstream of MPG1 (see above) and catalyzes the conversion of D-mannose-6-phosphate to D-mannose-1-phosphate, which is required for the initial steps of protein glycosylation.

A second form of CDGS, designated CDGS type 1B, has also been described (Niehues, R. et al. (1998) J. Clin. Invest. 101:1414-1420). Psychomotor dysfunction and mental retardation are not present in this disease; instead, CDGS type 1B is a gastrointestinal disorder characterized by protein-losing enteropathy, severe hypoglycemia, vomiting, diarrhea, and congenital hepatic fibrosis. Nonetheless, some patients who are affected with CDGS type 1B suffer from thrombosis and life-threatening bleeding. A deficiency of phosphomannose isomerase (PMI) was identified as the most likely cause of this syndrome. Most symptoms can be controlled with dietary mannose supplements (Niehues et al., *supra*; Freeze and Aebi, *supra*). This form of CDGS is the first inherited disorder in human metabolism that shows a decrease in available mannose.

Defects in glucosyltransferase function also play an important role in some human diseases. Galactosyltransferases may be involved in autoimmune/inflammatory disorders as many humans with autoimmune thyroid disorders have high levels of circulating antibodies directed against the enzymatic product of α1,3-galactosyltransferase (Etienne-Decerf, J. et al. (1987) Acta Endocrinol. 115:67-74). An aberrantly-cleaved, soluble β1,4-galactosyltransferase is secreted by a human ovarian cancer cell line (Uejima, T. et al. (1992) Cancer Res. 52:6158-6163). β1,4-GalT-deficient transgenic mice exhibited growth retardation in one experiment (Asano, M. et al. (1997) EMBO J. 16:1850-1857), while targeted inactivation of the mouse α1,4-GalT in another study was usually lethal (Furukawa, K. et al. (1999) Biochim. Biophys. Acta. 1473:54-66). In a third study, the constitutive overexpression of an α1,3-galactosyltransferase in transgenic mice led to the increased secretion of proteins in the urine, low body weight, partial damage to hair growth, and early death (Ikematsu, S. et al. (1999) Glycoconj. J. 1999 16:73-76). Galactosyltransferases have also been implicated in the regulation of cellular growth, development, and differentiation and may play an important role in embryogenesis as well as tumor development. Secreted galactosyltransferases, derived in some cases from proteolytic cleavage of membrane-bound forms, may trigger cell surface receptors by binding their bound carbohydrates or may modify carbohydrates on cell surface molecules in a regulated fashion. Extracellular carbohydrate moieties are developmentally regulated and are likely involved in the regulation of cell migration (Shur, B. et al. (1984) Mol. Cell. Biochem. 61:143-158; Paulson, J. and Colley, K. (1989) J. Biol. Chem. 264:17615-17618). The expression of β1,6-GlcNAc-bearing N-linked glycoproteins has been used as a marker of tumor progression in human breast and colon cancer, and astrocytes from human glioma specimens were found to contain increased levels of these type of glycoproteins compared to astrocytes from normal individuals (Yamamoto, H. et al. (2000) Cancer Res. 2000 60:134-142). These observations suggest that the dysfunction of another isoform of a glucosyltransferase, a β1,6-GlcNAcT, may also play a role in tumor formation or invasivity.

Sialyltransferases have also been implicated in human disease. Elevated levels of 2,6-sialyltransferase (but not 2,3-sialyltransferase) are detected in human choriocarcinoma tissues, apparently the result of upregulation at the transcriptional level (Fukushima, K. (1998) Cancer Res. 58:4301-4306). Transient transfection of 2,6-sialyltransferase into human, tumorigenic, glioma cell line, reduces the invasivity of the cells (Yamamoto, H. (1997) J. Neurochem. 68:2566-2576). Chronic alcohol (ethanol) consumption causes a decrease in Gal-β1,4-GlcNAc-α2,6-sialyltransferase (α2,6-ST) activity in the livers of rats who obtained at least one-third of their calories from alcohol for a period of one month or longer. Liver α2,6-ST activity returned to normal after a week of abstinence from alcohol consumption. Based on the results of nuclear run-on assays and mRNA stability assays, the reduction in α2,6-ST activity was the result of a 50% decrease in the half-life of

α2,6-ST mRNA (Rao, M. (1999) Metabolism 48:797-803). A significant decrease in the plasma α2,6-sialyltransferase activity was also observed in a group of individuals suffering from clinical depression. This particular form of depression was attributed to a change in glucocorticoid receptor (GR) functionality. These findings suggested that α2,6-ST enzyme or activity levels may be a contributing factor in clinical depression or at least a useful biochemical marker of cortisol receptor dysfunction (Maguire, T. et al. (1997) Biological Psychiatry 41:1131-1136).

Additional human diseases that involve defects in glycosylation, and the enzyme deficiencies that cause them, include (i) aspartylglycosaminuria, an aspartylglycosaminidase deficiency that causes mental retardation, (ii) GM_1 and GM_2 gangliosidosis, β-galactosidase and β-N-acetylhexosaminidase deficiencies, respectively, that cause glycolipid storage diseases, (iii) α-mannosidosis and β-mannosidosis, caused by a deficiency of α-mannosidase or β-mannosidase, respectively, that cause neurological dysfunction, and (iv) sialidosis, caused by a neuraminidase deficiency, characterized by hepatosplenomegaly as well as impaired neural development.

Expression Profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for examining which genes are tissue specific, carrying out housekeeping functions, parts of a signaling cascade, or specifically related to a particular genetic predisposition, condition, disease, or disorder.

The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of disease. For example, both the levels and sequences expressed in tissues from subjects with colon cancer may be compared with the levels and sequences expressed in normal tissue.

Colon cancer is causally related to both genes and the environment. Several molecular pathways have been linked to the development of colon cancer, and the expression of key genes in any of these pathways may be lost by inherited or acquired mutation or by hypermethylation. There is a particular need to identify genes for which changes in expression may provide an early indicator of colon cancer or a predisposition for the development of colon cancer.

For example, it is well known that abnormal patterns of DNA methylation occur consistently in human tumors and include, simultaneously, widespread genomic hypomethylation and localized areas of increased methylation. In colon cancer in particular, it has been found that these changes occur early in tumor progression such as in premalignant polyps that precede colon cancer. Indeed, DNA methyltransferase, the enzyme that performs DNA methylation, is significantly increased in histologically normal mucosa from patients with colon cancer or the benign polyps that precede

cancer, and this increase continues during the progression of colonic neoplasms (Wafik, S. et al. (1991) Proc. Natl. Acad. Sci. USA 88:3470-3474). Increased DNA methylation occurs in G+C rich areas of genomic DNA termed "CpG islands" that are important for maintenance of an "open" transcriptional conformation around genes, and hypermethylation of these regions results in a "closed" conformation that silences gene transcription. It has been suggested that the silencing or downregulation of differentiation genes by such abnormal methylation of CpG islands may prevent differentiation in immortalized cells (Antequera, F. et al. (1990) Cell 62:503-514).

Familial Adenomatous Polyposis (FAP) is a rare autosomal dominant syndrome that precedes colon cancer and is caused by an inherited mutation in the adenomatous polyposis coli (APC) gene. FAP is characterized by the early development of multiple colorectal adenomas that progress to cancer at a mean age of 44 years. The APC gene is a part of the APC-β-catenin-Tcf (T-cell factor) pathway. Impairment of this pathway results in the loss of orderly replication, adhesion, and migration of colonic epithelial cells that results in the growth of polyps. A series of other genetic changes follow activation of the APC-β-catenin-Tcf pathway and accompanies the transition from normal colonic mucosa to metastatic carcinoma. These changes include mutation of the K-ras proto-oncogene, changes in methylation patterns, and mutation or loss of the tumor suppressor genes p53 and Smad4/ DPC4. While the inheritance of a mutated APC gene is a rare event, the loss or mutation of APC and the consequent effects on the APC-β-catenin-Tcf pathway is believed to be central to the majority of colon cancers in the general population.

Hereditary nonpolyposis colorectal cancer (HNPCC) is another inherited autosomal dominant syndrome with a less well defined phenotype than FAP. HNPCC, which accounts for about 2% of colorectal cancer cases, is distinguished by the tendency to early onset of cancer and the development of other cancers, particularly those involving the endometrium, urinary tract, stomach and biliary system. HNPCC results from the mutation of one or more genes in the DNA mismatch-repair (MMR) pathway. Mutations in two human MMR genes, MSH2 and MLH1, are found in a large majority of HNPCC families identified to date. The DNA MMR pathway identifies and repairs errors that result from the activity of DNA polymerase during replication. Furthermore, loss of MMR activity contributes to cancer progression through accumulation of other gene mutations and deletions, such as loss of the BAX gene which controls apoptosis, and the TGF-β receptor II gene which controls cell growth. Because of the potential for irreparable damage to DNA in an individual with a DNA MMR defect, progression to carcinoma is more rapid than usual.

Although ulcerative colitis is a minor contributor to colon cancer, affected individuals have about a 20-fold increase in risk for developing cancer. Progression is characterized by loss of the p53 gene which may occur early, appearing even in histologically normal tissue. The progression of the disease from ulcerative colitis to dysplasia/carcinoma without an intermediate polyp state suggests a

high degree of mutagenic activity resulting from the exposure of proliferating cells in the colonic mucosa to the colonic contents.

Almost all colon cancers arise from cells in which the estrogen receptor (ER) gene has been silenced. The silencing of ER gene transcription is age related and linked to hypermethylation of the ER gene (Issa, J.P. et al. (1994) Nat. Genet. 7:536-540). Introduction of an exogenous ER gene into cultured colon carcinoma cells results in marked growth suppression. The connection between loss of the ER protein in colonic epithelial cells and the consequent development of cancer has not been established.

Clearly there are a number of genetic alterations associated with colon cancer and with the development and progression of the disease, particularly the downregulation or deletion of genes, that potentially provide early indicators of cancer development, and which may also be used to monitor disease progression or provide possible therapeutic targets. The specific genes affected in a given case of colon cancer depend on the molecular progression of the disease. Identification of additional genes associated with colon cancer and the precancerous state would provide more reliable diagnostic patterns associated with the development and progression of the disease.

The discovery of new carbohydrate-associated proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of carbohydrate metabolism, cell proliferative, autoimmune/inflammatory, reproductive, genetic, transport, and neurological disorders and cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of carbohydrate-associated proteins.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, carbohydrate-associated proteins, referred to collectively as "CHOP" and individually as "CHOP-1," "CHOP-2," "CHOP-3," "CHOP-4," "CHOP-5," "CHOP-6," "CHOP-7," "CHOP-8," "CHOP-9," and "CHOP-10." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-10.

The invention further provides an isolated polynucleotide encoding a polypeptide selected

from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-10. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:11-20.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an

immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid

sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional CHOP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CHOP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional CHOP, comprising

administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide

comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the SEQ ID NO:, Incyte ID, and length of the assembled polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold

parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"CHOP" refers to the amino acid sequences of substantially purified CHOP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CHOP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CHOP either by directly interacting with CHOP or by acting on components of the biological pathway in which CHOP participates.

An "allelic variant" is an alternative form of the gene encoding CHOP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times

in a given sequence.

"Altered" nucleic acid sequences encoding CHOP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CHOP or a polypeptide with at least one functional characteristic of CHOP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CHOP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CHOP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CHOP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CHOP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of CHOP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CHOP either by directly interacting with CHOP or by acting on components of the biological pathway in which CHOP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CHOP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the

translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and Gold, L. (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense

molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic CHOP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CHOP or fragments of CHOP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys

Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of CHOP or the polynucleotide encoding CHOP which is

identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:11-20 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:11-20, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:11-20 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:11-20 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:11-20 and the region of SEQ ID NO:11-20 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-10 is encoded by a fragment of SEQ ID NO:11-20. A fragment of SEQ ID NO:1-10 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-10. For example, a fragment of SEQ ID NO:1-10 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-10. The precise length of a fragment of SEQ ID NO:1-10 and the region of SEQ ID NO:1-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins and Sharp, 1989 (CABIOS 5:151-153) and in Higgins et al., 1992 (CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length

supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be

used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook et al., 1989 (Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions

will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CHOP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of CHOP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of CHOP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CHOP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where

necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an CHOP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of CHOP.

"Probe" refers to nucleic acid sequences encoding CHOP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989 (Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY), Ausubel et al., 1987 (Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY) and Innis et al., 1990 (PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer

selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid,

amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing CHOP, nucleic acids encoding CHOP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or

viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In one alternative, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al., 1989 (*supra*).

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a

propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human carbohydrate-associated proteins (CHOP), the polynucleotides encoding CHOP, and the use of these compositions for the diagnosis, treatment, or prevention of carbohydrate metabolism, cell proliferative, autoimmune/inflammatory, reproductive, genetic, transport, and neurological disorders and cancer.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s), along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the

MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are carbohydrate-associated proteins. For example, SEQ ID NO:1 is 54% identical, from residue Q170 to residue W909, to a murine α-glucosidase II, alpha subunit (GenBank ID g2104689) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.4e-258, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a glycosyl hydrolase family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:1 is a glycosyl hydrolase. In an alternative example, SEQ ID NO:4 is 55% identical, from residue E20 to residue T276, to bull α1,3-galactosyltransferase (GenBank ID g163124) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.8e-77, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from further BLAST analyses provide corroborative evidence that SEQ ID NO:4 is an α1,3-galactosyltransferase. In another example, SEQ ID NO:7 is 77% identical, from residue M2 to residue F402, to mouse β1,6-N-acetylglucosaminyltransferase B (GenBank ID g9650954) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.1e-177, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 also contains a core-2/I-branching enzyme domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from further BLAST analyses provide corroborative evidence that SEQ ID NO:7 is a β1,6-N-acetylglucosaminyltrasnferase B. In an alternative example, SEQ ID NO:9 is 81% identical, from residue M8 to residue Q200, to human intelectin (GenBank ID g8096221) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.1e-83, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 is an intelectin, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:9 also contains a fibrinogen domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST analysis of the PRODOM database provide further corroborative evidence that SEQ ID NO:9 is a lectin. In another example,

SEQ ID NO:10 is 36% identical, from residue T35 to residue E387, to human cargo selection protein (mannose-6-phosphate receptor binding protein) (GenBank ID g14043157) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.1e-51, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 also has homology to proteins that are localized to the endosomal vesicles, which mediate the transport of mannose-6-phosphate receptors from endosomes to the Golgi apparatus, and are endosome/endosomal vesicle cytoplasmic tail-interacting protein of 47 kDa and perilipin proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:10 also contains a perilipin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST analysis against the BLAST_PRODOM database provides further corroborative evidence that SEQ ID NO:10 is a homolog of perilipin. SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:8 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-10 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:11-20 or that distinguish between SEQ ID NO:11-20 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (i.e., those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (i.e., those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in

column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which

were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses CHOP variants. A preferred CHOP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CHOP amino acid sequence, and which contains at least one functional or structural characteristic of CHOP.

The invention also encompasses polynucleotides which encode CHOP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20, which encodes CHOP. The polynucleotide sequences of SEQ ID NO:11-20, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding CHOP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CHOP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:11-20. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CHOP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding CHOP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding CHOP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding CHOP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding CHOP. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CHOP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CHOP, some bearing minimal

similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CHOP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CHOP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CHOP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CHOP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CHOP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CHOP and CHOP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CHOP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:11-20 and fragments thereof under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied

Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A.(1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acid sequences encoding CHOP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate

software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CHOP may be cloned in recombinant DNA molecules that direct expression of CHOP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CHOP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CHOP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CHOP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding CHOP may be synthesized, in whole or in part, using chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, CHOP

itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of CHOP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active CHOP, the nucleotide sequences encoding CHOP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CHOP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CHOP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CHOP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CHOP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook et al., *supra*, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16).

A variety of expression vector/host systems may be utilized to contain and express sequences

encoding CHOP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook et al., *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CHOP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CHOP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of sequences encoding CHOP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke and Schuster, *supra*). When large quantities of CHOP are needed, e.g. for the production of antibodies, vectors which direct high level expression of CHOP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CHOP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation (Ausubel et al., 1995, *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994)

Bio/Technology 12:181-184).

Plant systems may also be used for expression of CHOP. Transcription of sequences encoding CHOP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CHOP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CHOP in host cells (Logan and Shenk, *supra*). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington et al., *supra*).

For long term production of recombinant proteins in mammalian systems, stable expression of CHOP in cell lines is preferred. For example, sequences encoding CHOP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or

herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CHOP is inserted within a marker gene sequence, transformed cells containing sequences encoding CHOP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CHOP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CHOP and that express CHOP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CHOP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CHOP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled

hybridization or PCR probes for detecting sequences related to polynucleotides encoding CHOP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CHOP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CHOP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CHOP may be designed to contain signal sequences which direct secretion of CHOP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CHOP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CHOP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CHOP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and

metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CHOP encoding sequence and the heterologous protein sequence, so that CHOP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al., 1995 (*supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CHOP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

CHOP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to CHOP. At least one and up to a plurality of test compounds may be screened for specific binding to CHOP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules. In one embodiment, the compound thus identified is closely related to the natural ligand of CHOP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified is a natural ligand of a receptor CHOP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, the compound can be closely related to the natural receptor to which CHOP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for CHOP which is capable of propagating a signal, or a decoy receptor for CHOP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Immunex Corp., Seattle WA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit CHOP involves producing appropriate cells which express CHOP, either as a secreted protein or on

the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing CHOP or cell membrane fractions which contain CHOP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either CHOP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with CHOP, either in solution or affixed to a solid support, and detecting the binding of CHOP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

CHOP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of CHOP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for CHOP activity, wherein CHOP is combined with at least one test compound, and the activity of CHOP in the presence of a test compound is compared with the activity of CHOP in the absence of the test compound. A change in the activity of CHOP in the presence of the test compound is indicative of a compound that modulates the activity of CHOP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising CHOP under conditions suitable for CHOP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CHOP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding CHOP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem

(ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding CHOP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding CHOP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding CHOP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress CHOP, e.g., by secreting CHOP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CHOP and carbohydrate-associated proteins. In addition, examples of tissues expressing CHOP can be found in Table 6 and can also be found in Example XI. Therefore, CHOP appears to play a role in carbohydrate metabolism, cell proliferative, autoimmune/inflammatory, reproductive, genetic, transport, and neurological disorders and cancer. In the treatment of disorders associated with increased CHOP expression or activity, it is desirable to decrease the expression or activity of CHOP. In the treatment of disorders associated with decreased CHOP expression or activity, it is desirable to increase the expression or activity of CHOP.

Therefore, in one embodiment, CHOP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CHOP. Examples of such disorders include, but are not limited to, a carbohydrate metabolic disorder, such as congenital type II dyserythropoietic anemia, diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, fructose-1,6-diphosphatase deficiency, galactosemia, glucagonoma, hereditary fructose intolerance, hypoglycemia, mannosidosis, neuraminidase deficiency, obesity, galactose epimerase deficiency, glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, carbohydrate-deficient glycoprotein syndromes (CDGS types 1A and 1B), autoimmune thyroid disorders, aspartylglycosaminuria, GM_1 gangliosidosis, GM_2 gangliosidosis, β-galactosidase β-N-acetylhexosaminidase deficiency, glycolipid storage diseases, neurological dysfunction, sialidosis, hepatosplenomegaly, and inherited abnormalities of pyruvate metabolism; a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency,

delayed puperty, retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumours; a genetic disorder, such as adrenoleukodystrophy, Alport's syndrome, choroideremia, Duchenne and Becker muscular dystrophy, Down syndrome, cystic fibrosis, chronic granulomatous disease, Gaucher's disease, Huntington's chorea, Marfan's syndrome, muscular dystrophy, myotonic dystrophy, pycnodysostosis, Refsum's syndrome, retinoblastoma, sickle cell anemia, thalassemia, Werner syndrome, von Willebrand's disease, Wilms' tumor, Zellweger syndrome, peroxisomal acyl-CoA oxidase deficiency, peroxisomal thiolase deficiency, peroxisomal bifunctional protein deficiency, mitochondrial carnitine palmitoyl transferase and carnitine deficiency, mitochondrial very-long-chain acyl-CoA dehydrogenase deficiency, mitochondrial medium-chain acyl-CoA dehydrogenase deficiency, mitochondrial short-chain acyl-CoA dehydrogenase deficiency, mitochondrial electron transport flavoprotein and electron transport flavoprotein:ubiquinone oxidoreductase deficiency, mitochondrial trifunctional protein deficiency, and mitochondrial short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency; a transport disorder, such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, Bell's palsy, Charcot-Marie Tooth disease, diabetes insipidus, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, dystonias, peripheral neuropathy; cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis; neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia; and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as ischemic cerebrovascular disease, stroke, cerebral neoplasms, Pick's disease, Huntington's disease, and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-

Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and seasonal affective disorder (SAD), akathesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, postherpetic neuralgia, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing CHOP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CHOP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified CHOP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CHOP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CHOP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CHOP including, but not limited to, those listed above.

In a further embodiment, an antagonist of CHOP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CHOP. Examples of such disorders include, but are not limited to, those carbohydrate metabolism, cell proliferative, autoimmune/inflammatory, reproductive, genetic, transport, and neurological disorders and cancer, described above. In one aspect, an antibody which specifically binds CHOP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CHOP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CHOP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CHOP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary

sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CHOP may be produced using methods which are generally known in the art. In particular, purified CHOP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CHOP. Antibodies to CHOP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with CHOP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CHOP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of CHOP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CHOP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P.

et al. (1984) Mol. Cell. Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CHOP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for CHOP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CHOP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CHOP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CHOP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of CHOP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CHOP epitopes, represents the average affinity, or avidity, of the antibodies for CHOP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular CHOP epitope, represents a true measure of affinity. High-affinity antibody preparations

with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the CHOP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CHOP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CHOP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*, and Coligan et al., *supra*).

In another embodiment of the invention, the polynucleotides encoding CHOP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding CHOP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CHOP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; Scanlon, K.J. et al. (1995) 9(13):1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736).

In another embodiment of the invention, polynucleotides encoding CHOP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-

linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in CHOP expression or regulation causes disease, the expression of CHOP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in CHOP are treated by constructing mammalian expression vectors encoding CHOP and introducing these vectors by mechanical means into CHOP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of CHOP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CHOP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter

(Rossi and Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding CHOP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Nëumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to CHOP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding CHOP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding CHOP to cells which have one or more genetic abnormalities with respect to the expression of CHOP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas

(Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi et al., 1999 (Annu. Rev. Nutr. 19:511-544) and Verma and Somia, 1997 (Nature 18:389:239-242), both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CHOP to target cells which have one or more genetic abnormalities with respect to the expression of CHOP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CHOP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins et al., 1999 (J. Virol. 73:519-532) and Xu et al., 1994 (Dev. Biol. 163:152-161), hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding CHOP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for CHOP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of CHOP-coding RNAs and the synthesis of high levels of CHOP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN)

indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of CHOP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994), in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CHOP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding CHOP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines,

cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding CHOP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased CHOP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding CHOP may be therapeutically useful, and in the treatment of disorders associated with decreased CHOP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding CHOP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CHOP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding CHOP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding CHOP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without

exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of CHOP, antibodies to CHOP, and mimetics, agonists, antagonists, or inhibitors of CHOP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the

lung have enabled the practical delivery of drugs such as insulin to blood circulation (Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising CHOP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CHOP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CHOP or fragments thereof, antibodies of CHOP, and agonists, antagonists or inhibitors of CHOP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and

response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind CHOP may be used for the diagnosis of disorders characterized by expression of CHOP, or in assays to monitor patients being treated with CHOP or agonists, antagonists, or inhibitors of CHOP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CHOP include methods which utilize the antibody and a label to detect CHOP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CHOP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CHOP expression. Normal or standard values for CHOP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to CHOP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CHOP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding CHOP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CHOP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CHOP, and to monitor regulation of CHOP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CHOP or closely related molecules may be used to identify nucleic acid sequences which encode CHOP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5'regulatory region, or from a less specific region, e.g., a

conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CHOP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CHOP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:11-20 or from genomic sequences including promoters, enhancers, and introns of the CHOP gene.

Means for producing specific hybridization probes for DNAs encoding CHOP include the cloning of polynucleotide sequences encoding CHOP or CHOP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CHOP may be used for the diagnosis of disorders associated with expression of CHOP. Examples of such disorders include, but are not limited to, a carbohydrate metabolic disorder, such as congenital type II dyserythropoietic anemia, diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, fructose-1,6-diphosphatase deficiency, galactosemia, glucagonoma, hereditary fructose intolerance, hypoglycemia, mannosidosis, neuraminidase deficiency, obesity, galactose epimerase deficiency, glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, carbohydrate-deficient glycoprotein syndromes (CDGS types 1A and 1B), autoimmune thyroid disorders, aspartylglycosaminuria, GM_1 gangliosidosis, GM_2 gangliosidosis, β-galactosidase β-N-acetylhexosaminidase deficiency, glycolipid storage diseases, neurological dysfunction, sialidosis, hepatosplenomegaly, and inherited abnormalities of pyruvate metabolism; a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis,

myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puperty, retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumours; a genetic disorder, such as adrenoleukodystrophy, Alport's syndrome, choroideremia, Duchenne and Becker muscular dystrophy, Down syndrome, cystic fibrosis, chronic granulomatous disease, Gaucher's disease, Huntington's chorea, Marfan's syndrome, muscular dystrophy, myotonic dystrophy, pycnodysostosis, Refsum's syndrome, retinoblastoma, sickle cell anemia, thalassemia, Werner syndrome, von Willebrand's disease, Wilms' tumor, Zellweger syndrome, peroxisomal acyl-CoA oxidase deficiency, peroxisomal thiolase deficiency, peroxisomal bifunctional protein deficiency, mitochondrial carnitine palmitoyl transferase and carnitine deficiency, mitochondrial very-long-chain acyl-CoA dehydrogenase deficiency, mitochondrial medium-chain acyl-CoA dehydrogenase deficiency, mitochondrial short-chain acyl-CoA dehydrogenase deficiency, mitochondrial electron transport flavoprotein and electron transport flavoprotein:ubiquinone oxidoreductase deficiency, mitochondrial trifunctional protein deficiency, and mitochondrial short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency; a transport disorder, such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, Bell's palsy, Charcot-Marie Tooth disease, diabetes insipidus, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, dystonias, peripheral neuropathy; cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis; neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression,

epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia; and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as ischemic cerebrovascular disease, stroke, cerebral neoplasms, Pick's disease, Huntington's disease, and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and seasonal affective disorder (SAD), akathesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, postherpetic neuralgia, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding CHOP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CHOP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CHOP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CHOP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a

suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CHOP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CHOP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CHOP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CHOP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding CHOP, or a fragment of a polynucleotide complementary to the polynucleotide encoding CHOP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding CHOP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are

substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding CHOP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of CHOP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is

presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, CHOP, fragments of CHOP, or antibodies specific for CHOP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed

molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical

density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for CHOP to quantify the levels of CHOP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are

incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach (M. Schena, ed. (1999) Oxford University Press, London), hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding CHOP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington et al., *supra*; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995), in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CHOP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse,

may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CHOP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CHOP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CHOP, or fragments thereof, and washed. Bound CHOP is then detected by methods well known in the art. Purified CHOP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CHOP specifically compete with a test compound for binding CHOP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CHOP.

In additional embodiments, the nucleotide sequences which encode CHOP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/293,768, U.S. Ser. No. 60/309,548, U.S. Ser. No. 60/314,400, U.S. Ser. No. 60/343,706, and U.S. Ser. No. 60/337,999, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel, *supra*, units 5.1-6.6). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid,

QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002)

Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:11-20. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative carbohydrate-associated proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg).

Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode carbohydrate-associated proteins, the encoded polypeptides were analyzed by querying against PFAM models for carbohydrate-associated proteins. Potential carbohydrate-associated proteins were also identified by homology to Incyte cDNA sequences that had been annotated as carbohydrate-associated proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants.

Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of CHOP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:11-20 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:11-20 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation

hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al., *supra*, ch. 7; Ausubel et al., 1995, *supra*, ch. 4 and 16).

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \times \text{minimum } \{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding CHOP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is

classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding CHOP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of CHOP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar,

Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in CHOP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:11-20 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice

variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:11-20 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc.

(CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibited at least about a two-fold change in expression, a signal-to-background

ratio of at least 2.5, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

Expression

For example, SEQ ID NO:19 showed differential expression in colon tissue from patients with colon cancer compared to matched microscopically normal tissue from the same donors as determined by microarray analysis. The expression of SEQ ID NO:19 was decreased at least two-fold in cancerous colon tissue. Therefore, SEQ ID NO:19 is useful in disease staging and diagnostic assays for cell proliferative disorders, including colon cancer.

XII. Complementary Polynucleotides

Sequences complementary to the CHOP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CHOP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CHOP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CHOP-encoding transcript.

XIII. Expression of CHOP

Expression and purification of CHOP is achieved using bacterial or virus-based expression systems. For expression of CHOP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CHOP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CHOP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CHOP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard et al., *supra*; Sandig et al., *supra*).

In most expression systems, CHOP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from CHOP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al., 1995 (*supra*, ch. 10 and 16). Purified CHOP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functional Assays

CHOP function is assessed by expressing the sequences encoding CHOP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, 1994 (Flow Cytometry, Oxford, New York NY).

The influence of CHOP on gene expression can be assessed using highly purified populations

of cells transfected with sequences encoding CHOP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CHOP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of CHOP Specific Antibodies

CHOP substantially purified using polyacrylamide gel electrophoresis (PAGE; Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the CHOP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., 1995, *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., 1995, *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CHOP activity by, for example, binding the peptide or CHOP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring CHOP Using Specific Antibodies

Naturally occurring or recombinant CHOP is substantially purified by immunoaffinity chromatography using antibodies specific for CHOP. An immunoaffinity column is constructed by covalently coupling anti-CHOP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CHOP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CHOP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CHOP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CHOP is collected.

XVII. Identification of Molecules Which Interact with CHOP

CHOP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CHOP, washed, and any wells with labeled CHOP complex are assayed. Data obtained using different concentrations of CHOP are used to calculate values for the number, affinity, and association of CHOP with the candidate molecules.

Alternatively, molecules interacting with CHOP are analyzed using the yeast two-hybrid system as described in Fields and Song, 1989 (Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

CHOP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of CHOP Activity

Galactosyltransferase Activity Assay #1

β1,3-galactosyltransferase and β1,4-galactosyltransferase activity of CHOP is determined by measuring the transfer of galactose from UDP-galactose to a GlcNAc-terminated oligosaccharide chain in a radioactive assay (Kolbinger, F. et al. (1998) J. Biol. Chem. 273:58-65; Hennet, T. et al. (1998) J. Biol. Chem. 273:58-65). An aliquot of CHOP is incubated with 14 μl of assay stock solution (180 mM sodium cacodylate, pH 6.5, 1 mg/ml bovine serum albumin, 0.26 mM UDP-galactose, 2 μl of UDP-[^{3}H]galactose, 1 μl of $MnCl_2$ (500 mM), and 2.5 μl of GlcNAcβO-$(CH_2)_8$-CO_2Me (37 mg/ml in dimethyl sulfoxide)) for approximately 1 hr at 37°C. The reaction is quenched by the addition of 1 ml of water and loaded on a C18 Sep-Pak cartridge (Waters). The column is washed twice with 5 ml of water to remove unreacted UDP-[^{3}H]galactose. The reaction product, [^{3}H]galactosylated GlcNAcβO-$(CH_2)_8$-CO_2Me, remains bound to the column during the water washes and is eluted with 5 ml of methanol. The level of radioactivity in the eluted material is measured by liquid scintillation counting and is proportional to CHOP galactosyltransferase activity in the sample.

Galactosyltransferase Activity Assay # 2

In the alternative, β1,4-galactosyltransferase activity of CHOP is determined by quantitating the transfer of [^{14}C]galactose from UDP-[^{14}C]Gal to ovalbumin. The approximately 50 μl reaction contains 50 mM HEPES (pH 7.35), 10 mM $MnCl_2$, 1.5 mg of ovalbumin, 50 mM NaCl, 5 μl UDP-[^{14}C]galactose (25 nCi), and 5 μl of CHOP. The assay is incubated at 60 °C for 30 minutes and terminated by the addition of ice-cold 2.5% phosphotungstic acid (w/v) in 1 M HCl. Unincorporated UDP-[^{14}C]Gal is separated by filtration through Whatman GF/C glass fiber filters. The filters are washed once with 2.5% phosphotungstic acid (w/v) and then rinsed with ice-cold ethanol. The filters

are dried, and radioactivity is determined using a scintillation counter. The amount of radioactivity is proportional to the activity of CHOP (Verdon, B. and Berger, E. (1983) in Galactosyltransferase. Methods of Enzymatic Analysis, Bergmeyer, H. et al. (eds.), Vol. III, 3rd Ed., Verlag Chemie, Weinheim, Deerfield Beach, Basel, pp. 374-381; Tilo, S. et al. (1996) J. Biol. Chem. 271:3398-3405).

Sialyltransferase Activity Assay #1

In the alternative, sialyltransferase activity of CHOP is assayed as follows. The sialyltransferase acceptors used in the assay are derived from aminophenylglycosides reacted with 6(5-fluorescein-carboxamido)-hexanoic acid succimidyl ester (FCHASE). Briefly, 10 mg *p*-aminophenylglycoside is dissolved in 0.5 ml of 0.2 M triethylamine acetate buffer, pH 8.2. FCHASE was dissolved in 0.5 ml of methanol and added to the aminophenylglycoside solution. The mixture is stirred in the dark for about 3 h at room temperature, lyophilized, resuspended in approximately 200 μl of 50% acetonitrile, and spotted on a silica thin-layer chromatography (TLC) plate which is developed with an ethyl acetate/methanol/water/acetic acid solvent system. Following air drying in a fume hood, the yellow product is scraped, eluted with distilled water, concentrated, desalted, and bound to a Sep-Pak C18 reverse phase cartridge. After washing the cartridge with several volumes of water, the product is eluted in 50% acetonitrile and quantitated by well-known spectrophotometric methods. Following preparation of the sialyltransferase acceptors, enzyme reactions are performed at 37 °C in 20 μl volumes in a buffer consisting of 50 mM MES (pH 6.0), 10 mM $MnCl_2$, with 0.2 or 1.0 mM labeled acceptor, 0.2 mM CMP-Neu5Ac donor, and various amounts of CHOP. The reaction is terminated by diluting the reaction with 10 mM NaOH prior to analysis by capillary electrophoresis. Capillary electrophoresis (CE) is performed using an Argon-ion laser-induced fluorescence detector (excitation = 488 nm; emission = 520 nm). The product peak, consisting of FCHASE-2,3-sialyl-N-acetyllactosamine, is identified and quantitated and is proportional to the sialyltransferase activity in the sample (Wakarchuk, W. et al. (1996) J. Biol. Chem. 271:19166-19173; Gilbert, M. (1996) J. Biol. Chem. 271:28271-28276).

Sialyltransferase Activity Assay #2

Alternatively, sialyltransferase activity of CHOP is assayed as follows. Sialyltransferase assays are performed in a reaction mixture containing 10 mM $MgCl_2$, 0.3% Triton CF-54, 100 mM sodium cacodylate buffer (pH 6.0), 0.66 mM unlabeled CMP-Neu5Ac donor, 4,400 dpm/μl CMP-[^{14}C]Neu5Ac (tracer), the CHOP solution, and substrates in total volume of 20 to 50 μl. The reaction mixture is incubated at 37 °C for 2 h then terminated by addition of 500 μl of water. The products are isolated by C18 Sep-Pak cartridge and analyzed by thin layer chromatography (TLC) using the solvent system ethanol/pyridine/n-butanol/acetate/water (100:10:10:3:30) (Okajima, T. et al. (1999) J. Biol. Chem. 274:11479-11486). The radiolabeled products are visualized by standard autoradiography techniques familiar to persons skilled in the art and sialyltransferase activity of

CHOP is proportional to the signal on the autoradiogram.

Sialyltransferase Activity Assay #3

In the alternative, sialyltransferase activity of CHOP is assayed as follows. Human embryonic kidney cells (293) are stably transfected with a plasmid encoding the CHOP. The cells are grown to confluence in 225 cm^2 tissue culture flasks and harvested by scraping cells into phosphate-buffered saline (PBS). Cells are pelleted and resuspended in approximately 1 ml of 1% Triton X-100, 50 mM NaCl, 5 mM MnCl, and 25 mM MES (pH 6.0). The cell pellet is solubilized by repeated pipetting and vortexing. This crude homogenate is cleared by centrifugation at 1000 x g for 10 min and used directly as the enzyme source. The assay mixture consists of 50 μM CMP-Neu5Ac with 250,000 cpm of CMP-[^{14}C]Neu5Ac added as a tracer, 0.1% Triton CF-54, 20 mM cacodylate (pH 6.0) and 10 μl of CHOP-containing extract in a 30-μl reaction volume. Glycoprotein and glycolipid products are separated from CMP-Neu5Ac by gel filtration and the amount of label in the eluted fractions are quantitated to determine the relative amount of CHOP activity in the sample (Sjoberg, E. et al. (1996) J. Biol. Chem. 271:7450-7459).

O-glucosyltransferase Transferase Activity Assay

O-glucosyltransferase activity of CHOP is assayed as follows. CHOP preparations are diluted with cold desalt buffer (20 mM Tris pH 8.0, 20% glycerol, 0.02% NaN_3) immediately prior to use. Peptide substrates (Kreppel, L. et al. (1999) J. Biol. Chem. 274:32015-32022) are used as acceptors at a concentration of approximately 3 mM. The peptides in the reaction are separated from the reactants using a SP-Sephadex column. The modified and unmodified peptides are loaded onto a Sep-pak C18 cartridge. Unmodified peptides are eluted with 50 mM formic acid, 10 ml of 50 mM formic acid containing 0.5 M NaCl, and 10 ml of distilled H_2O. Modified peptides are eluted from the cartridge directly into scintillation vials using methanol. Enzyme activity is expressed in terms of micromoles of GlcNAc transferred per minute, which is proportional to the level of CHOP activity in the sample (Kreppel et al., *supra*).

Mannosidase Activity Assay

Mannosidase activity in CHOP is measured by its ability to release mannose from $Man_9(GlcNAc)_2$ oligosaccharide (Schweden, J. et al. (1986) Eur. J. Biochem. 157:563-570). CHOP, in 200 mM phosphate buffer, pH 6.5 and 1% Triton X-100, is mixed with [^{14}C]$(Man_9)(GlcNAc)_2$ (2-3 x 10^3 cpm) in a final volume of 30 μl at 37°C for 60 minutes. The reaction is terminated by the addition of 30 μl glacial acetic acid. The amount of liberated [^{14}C]mannose, analyzed by paper chromatography in 2-propanol/acetic acid/water (29/4/9, by volume), is proportional to the activity of CHOP in the starting sample.

Carbohydrate Binding Assay

CHOP activity is also demonstrated by the ability of CHOP to bind to β-galactoside sugars.

CHOP is applied to a lactosyl-Sepharose column, and the column is eluted with 0.1 M lactose. The presence of CHOP in the eluate is detected by sodium dodecyl sulfate polyacrylamide gel electrophoresis and indicates the ability of CHOP to bind β-galactoside sugars.

Hyaluronan Hydrolysis Assay

CHOP activity is also measurable by its ability to hydrolyze hyaluronan (HA) (Lepperdinger, *supra*). Radioactively labeled HA is immobilized on microtiter plates with the aid of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide and N-hydroxy-sulfosuccinimide. The radioactivity solubilized after incubation with CHOP is measured using a liquid scintillation counter and is proportional to the activity of CHOP in the starting sample.

Cellular Transformation Assay

Alternatively, CHOP activity is measured by its ability to regulate transformation of NIH3T3 mouse fibroblast cells. A cDNA encoding CHOP is subcloned into an appropriate eukaryotic expression vector. This construct is transfected into NIH3T3 cells using methods known in the art. Transfected cells are compared with non-transfected cells for the following quantifiable properties characteristic of oncogenically-transformed cells: high-density growth in culture associated with loss of contact inhibition; growth in suspension or in soft agar; reduced serum requirements; and ability to induce tumors when injected into immunodeficient mice. The activity of CHOP is proportional to the extent of transformation of NIH3T3 cells transfected with CHOP, compared to non-transfected cells.

XIX. Expression and purification of CHOP-associated α1,3-Galactosyltransferase activity

Soluble, catalytically-active CHOP with α1,3-galactosyltransferase (α1,3-GalT) activity is purified from culture media of nearly confluent human embryonic kidney (293) cells transfected with a plasmid expressing CHOP. The culture media is dialyzed overnight at 4 °C against 20 mM HEPES (pH 7.0) and 5 mM $MnCl_2$, with several changes of buffer and then applied to an affinity column of UDP-hexanolamine-Sepharose (4 μmol/ml resin). The column is washed with dialysis buffer and dialysis buffer containing 0.75 M NaCl until the A_{280} of the eluate reaches background levels. The bound enzyme is eluted with 5 mM UDP in dialysis buffer and concentrated using a Centricon-10 (Millipore). The homogeneity of the purified enzyme is determined by separation on a 10% SDS-PAGE gel (Kim, S. et al. (1997) J. Biol. Chem. 272:13622-13628).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7247177	1	7247177CD1	11	7247177CB1
5596327	2	5596327CD1	12	5596327CB1
4179651	3	4179651CD1	13	4179651CB1
7482109	4	7482109CD1	14	7482109CB1
2867618	5	2867618CD1	15	2867618CB1
7488348	6	7488348CD1	16	7488348CB1
5539166	7	5539166CD1	17	5539166CB1
7500341	8	7500341CD1	18	7500341CB1
7500846	9	7500846CD1	19	7500846CB1
3230921	10	3230921CD1	20	3230921CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	Annotation
1	7247177CD1	g2104689	1.4E-258	[Mus musculus] alpha-glucosidase II, alpha subunit Arendt, C.W. and Ostergaard, H.L. (1997) J. Biol. Chem.272:13117-13125.
2	5596327CD1	g2621120	1.9E-17	[Methanothermobacter thermautotrophicus] O-linked GlcNAc transferase Smith, D.R. et al. (1997) J. Bacteriol. 179:7135-7155.
3	4179651CD1	g14150450	0.0	[Rattus norvegicus] UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T9 Ten Hagen, K.G. et al. (2001) J. Biol. Chem. 276:17395-17404.
4	7482109CD1	g163124	5.8E-77	[Bos taurus] alpha-1,3-galactosyltransferase Joziasse, D.H. et al. (1989) J. Biol. Chem. 264:14290-14297.
5	2867618CD1	g2642187	4.4E-182	[Rattus norvegicus] endo-alpha-D-mannosidase Spiro, M.J. et al. (1997) J. Biol. Chem. 272:29356-29363.
6	7488348CD1	g31179	4.1E-243	[Homo sapiens] enolase
7	5539166CD1	g9650954	2.1E-177	[Mus musculus] beta-1,6-N-acetylglucosaminyltransferase B Chen, G.Y. et al. (2000) Glycobiology 10:1001-1011.
8	7500341CD1	g3790363	5.4E-30	[Homo sapiens] DPM2 Maeda, Y. et al. (1998) EMBO J. 17:4920-4929. Lennon, G. et al. (1996) Genomics 33:151-152.
		339872|DPM2	4.8E-31	[Homo sapiens][Regulatory subunit; Transferase] [Endoplasmic reticulum; Cytoplasmic] Dolichol phosphate mannose synthase regulatory subunit, regulates synthesis of dolichol phosphate mannose, which is a donor of mannosyl residues for N-linked oligosaccharide precursors and for the core of glycosylphosphatidylinositol anchors
9	7500846CD1	g8096221	2.1E-83	[Homo sapiens] intelectin Tsuji, S. et al. (2001) J. Biol. Chem. 276:23456-23463.
		585107|Itln	4.2E-80	[Mus musculus] Intelectin, expressed in small intestinal Paneth cells and may have a role in the defense against microorganisms
10	3230921CD1	g14043157	1.1E-51	[Homo sapiens] cargo selection protein (mannose-6-phosphate receptor binding protein) Diaz, E. and Pfeffer, S.R. (1998) Cell 93:433-443.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	Annotation
10 cont.		342866|TIP47	9.2E-53	[Homo sapiens][Endosome/Endosomal vesicles; Cytoplasmic] Tail-interacting protein of 47 kDa, protein that binds to the cytoplasmic domain of mannose-6-phosphate receptors and mediates their transport from endosomes to the Golgi apparatus; serum levels are elevated in cervical carcinoma and normal pregnancy Orsel, J. G., et al. (2000) Proc. Natl. Acad. Sci. USA 97:9047-9051.
		339060|ADFP	9.8E-33	[Homo sapiens][Endoplasmic reticulum; Cytoplasmic; Lipid particles] Adipose differentiation-related protein, a membrane-associated protein involved in fatty acid transport, a marker of lipid accumulation
		583559|Adfp	2.7E-26	[Mus musculus][Small molecule-binding protein][Cytoplasmic; Unspecified membrane] Adipose differentiation-related protein, a membrane-associated protein involved in fatty acid and cholesterol binding and long-chain fatty acid transport, a marker of lipid accumulation Tansey, J. T., et al. Proc. Natl. Acad. Sci. U.S.A. 98:6494-9. (2001)
		337032|PLIN	2.8E-16	[Homo sapiens] Perilipin, may be multiply phosphorylated by protein kinase A in adipocytes
		329340|Plin	3.5E-16	[Rattus norvegicus][Cytoplasmic; Lipid particles] Perilipin, regulates lipolysis and is multiply phosphorylated by protein kinase A in adipocytes

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7247177CD1	912	S10 S38 S56 S145 S427 S477 S577 S804 S809 S835 S844 T42 T51 T52 T103 T108 T180 T184 T290 T340 T650 T732 T759 T907 Y246 Y449	N286 N843	Glycosyl hydrolases family: P138-T875	HMMER_PFAM
					Transmembrane domain: W592-F620 N-terminus is cytosolic	TMAP
					Glycosyl hydrolases family BL00129: G353-D398, G464-Y490, R565-W592, L603-A646, R667-F702, T783-S820	BLIMPS_BLOCKS
					HYDROLASE GLYCOSIDASE GLYCOPROTEIN ALPHA-GLUCOSIDASE MALTASE PRECURSOR SIGNAL PROTEIN GLUCOSIDASE SUCRASE ISOMALTASE PD001543: W195-F518, Q523-T874	BLAST_PRODOM
					HYDROLASE GLYCOSIDASE ALPHA-GLUCOSIDASE GLYCOPROTEIN MALTASE SUCRASE ISOMALTASE PRECURSOR SIGNAL ACID INTESTINAL PD001716: G329-W472, H224-Y302	BLAST_PRODOM
					SUCRASE/ISOMALTASE DM01359|P10253|285-950: H537-L898 DM01359|P23739|278-931: Q523-A906 DM01359|P14410|268-925: N534-I905 DM01359|JC4624|199-864: E522-Q822	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1 cont.					Glycosyl hydrolase family 31 signature: F507-E514	MOTIFS
2	5596327CD1	691	S80 S137 S198 S291 S327 S510 S615 S658 S664 S674 S684 T99 T541 T561 T671 Y270		ATP synthase Alpha chain, C-terminal: R128-L142	HMMER_PFAM
					TPR Domain: I149-H182, V183-N216, P115-F148, P48-A81, R82-Q114	HMMER_PFAM
					Transmembrane domain: E575-Y591, P595-K617 N-terminus is non-cytosolic	TMAP
					MITOCHONDRIAL OUTER MEMBRANE PROTEIN, 70K DM04281|P23231|276-618: I17-Q180	BLAST_DOMO
3	4179651CD1	506	S21 S29 S40 S75 S90 S108 S141 S253 S319 S337 S443 S465 S470 T2 T114 T137 T409 T422 T437 T498 Y196 Y238 Y459	N27 N49 N496	Glycosyl transferase: S51-G236	HMMER_PFAM
					QXW lectin repeat: G417-D455, K456-T495, G364-N402	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3 cont.					N-ACETYLGALACTOSAMINYLTRANSFERASE TRANSFERASE POLYPEPTIDE ACETYLGALACTOSAMINYLTRANSFERASE UDP-GALNAC:POLYPEPTIDE GLYCOSYLTRANSFERASE PROTEIN UDP PROTEIN UDP N PD003162:I202-A362	BLAST_PRODOM
					do ACETYLGALACTOSAMINYL-TRANSFERASE; POLYPEPTIDE DM03891|Q07537|32-558: D3-F503 DM03891|P34678|37-600: D8-E493 DM03891|I37405|21-571: Y11-F492	BLAST_DOMO
4	7482109CD1	276	S3 S153 S260 T47 T84 T131 T139 T213 Y70	N74 N259	signal_cleavage: M1-V19	SPSCAN
					Signal Peptide: M1-R21, M1-Y22, M1-R24	HMMER
					TRANSFERASE ALPHA 1,3-GALACTOSYLTRANSFERASE N-ACETYLLACTOSAMINIDE GALACTOSYLTRANSFERASE UDP-GALACTOSE: BETA D-GALACTOSYL 1 GROUP 4-N-ACETYL D-GLUCOSAMINIDE GLYCOSYLTRANSFERASE PD010022:K41-T276	BLAST_PRODOM
					SIGNAL-ANCHOR TRANSMEM (GALACTOSYLTRANSFERASE FUCOSYLGLYCOPROTEIN ALPHA-1 ALPHA) DM07533|P14769|6-367: E20-T276 DM07533|P16442|16-353: R42-T276	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4 cont.					do GALACTOSYLTRANSFERASE; ALPHA-1; ACETYLLACTOSAMINIDE; DM08008|A56480|1-376: E20-T276 DM08008|I49698|1-371: R24-T276	BLAST_DOMO
5	2867618CD1	378	S80 S85 S206 S256 S280 S317 S336 S359 T125 T190 T325 T355 Y369		ENDO-ALPHA D-MANNOSIDASE PD141586: P3-S378	BLAST_PRODOM
6	7488348CD1	458	S263 S272 S306 S370 S392 S424 T27 T246 T341 T421 Y61 Y279	N76 N108	Enolase: S2-R456	HMMER_PFAM
					Enolase proteins BL00164: F403-A441, R35-D57, I153-K202, N229-K271, P302-D316, I331-N366	BLIMPS_BLOCKS
					Enolase signature: Q314-L362	PROFILESCAN
					Enolase signature PR00148: V38-E52, G113-G129, A173-R186, G335-R346, L361-L375, S392-V409	BLIMPS_PRINTS
					ENOLASE LYASE GLYCOLYSIS MAGNESIUM HYDROLYASE 2-PHOSPHOGLYCERATE DEHYDRATASE 2-PHOSPHO-D-GLYCERATE 2-PHOSPHO-D-GLYCERATE PD000902: M1-V293	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6 cont.					LYASE MAGNESIUM ENOLASE GLYCOLYSIS HYDROLYASE DEHYDRATASE 2-PHOSPHOGLYCERATE 2-PHOSPHO-D-GLYCERATE 2-PHOSPHO-D-GLYCERATE PD003216: D275-E440	BLAST_PRODOM
					ENOLASE LYASE GLYCOLYSIS MAGNESIUM HYDROLYASE 2-PHOSPHOGLYCERATE DEHYDRATASE 2-PHOSPHO-D-GLYCERATE 2-PHOSPHO-D-GLYCERATE PD000948: V408-R456	BLAST_PRODOM
					ENOLASE DM00487|S22071|2-455: S2-R456 DM00487|S52858|1-431: M1-P455 DM00487|JC1039|1-431: M1-P455 DM00487|P15429|1-430: I7-R456	BLAST_DOMO
					Enolase signature: L361-S374	MOTIFS
7	5539166CD1	402	S4 S89 S154 S158 S295 S329 T67 T68 T73 T87 T199 T247 Y206	N41 N316 N390	signal_cleavage: M1-N25	SPSCAN
					Core-2/I-Branching enzyme: T73-L382	HMMER_PFAM
					Cytosolic Domain: M1-K6 Transmembrane Domain: H7-W29 Non-cytosolic Domain: E30-F402	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7 cont.					PROTEIN BETA 1-TRANSFERASE GLYCOSYLTRANSFERASE ENZYME CORE 6-N-ACETYLGLUCOSAMINYLTRANSFERASE 3-GALACTOSYLOGLYCOSYL GLYCOPROTEIN 6-N-ACETYLGLUCOSAMINYLTRANSFERASE PD005410: L171-H344 PD003538: T73-D170	BLAST_PRODOM
					BETA 1-TRANSFERASE GLYCOSYLTRANSFERASE ENZYME CORE 6-N-ACETYLGLUCOSAMINYLTRANSFERASE 3-GALACTOSYLOGLYCOSYL GLYCOPROTEIN 6-N-ACETYLGLUCOSAMINYLTRANSFERASE BRANCHING PD011484: G345-F402	BLAST_PRODOM
					LUMENAL DOMAIN DM07544|Q06430|8-399: L9-F402 DM07544|Q02742|29-427: C74-R385	BLAST_DOMO
8	7500341CD1	109	T3 T73		signal_cleavage: M1-S67	SPSCAN
					Cytosolic domains: M1-V8, P69-D109 Transmembrane domains: V9-L31, P46-F68 Non-cytosolic domain: P32-L45	TMHMMER
9	7500846CD1	225	S30 S52 S62 S96 T80 T107 T142		signal_cleavage: M1-A26	SPSCAN
					Signal Peptide: M1-A27, M1-A28, M4-A26, M4-A28, M8-C24, M8-A26, M8-A28, M8-A29	HMMER
					Fibrinogen beta and gamma chains, C-terminus: L49-V94	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9 cont.					Cytosolic domain: M1-C12 Transmembrane domain: F13-L35 Non-cytosolic domain: S36-G225	TMHMMER
					LECTIN INTELECTIN CORTICAL GRANULE PRECURSOR GLYCOPROTEIN PD040823: R51-Q200	BLAST_PRODOM
10	3230921CD1	463	S2 S13 S117 S140 S187 S205 S273 S277 S304 T68 T169		signal_cleavage: M1-A38	SPSCAN
					Perilipin family: P10-V369	HMMER_PFAM
					PROTEIN PERILIPIN ADIPOSE DIFFERENTIATION RELATED ADRP MEMBRANE CARGO SELECTION TIP47 A/B PD018256: S148-E384, Q19-G128	BLAST_PRODOM

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
11/7247177CB1/ 3150	1-388, 1-394, 1-730, 1-737, 5-358, 5-423, 35-603, 220-394, 228-423, 239-978, 239-980, 285-389, 295-485, 295-905, 297-544, 311-418, 311-709, 454-719, 498-1352, 508-1352, 546-1352, 549-669, 549-751, 549-812, 550-1352, 568-1168, 580-1352, 585-1352, 589-1088, 589-1236, 619-1352, 681-1352, 790-1650, 790-1661, 821-1714, 872-1147, 876-1584, 885-1584, 956-1608, 971-1615, 1002-1621, 1063-1942, 1070-1961, 1075-1277, 1098-1704, 1099-1781, 1099-1864, 1184-1635, 1193-1893, 1224-1901, 1235-1955, 1236-1961, 1250-1955, 1315-1574, 1324-1955, 1345-1955, 1398-1967, 1457-1954, 1496-1721, 1498-1949, 1505-1952, 1505-1955, 1512-1595, 1518-1864, 1651-1864, 1786-2256, 1817-2103, 1827-2567, 1827-2584, 1827-2598, 1827-2614, 1827-2637, 1831-2622, 1865-2227, 1892-2474, 1894-2196, 1918-2032, 1923-2092, 1946-2324, 1975-2840, 1988-2586, 2188-2897, 2200-2897, 2205-2897, 2206-2897, 2208-2897, 2413-2526, 2417-2885, 2447-3062, 2489-3150, 2605-2951, 2641-2804, 2789-2932, 2802-2951, 2869-2951, 3113-3138
12/5596327CB1/ 3290	1-788, 303-1028, 315-1031, 591-798, 662-962, 683-960, 683-1394, 812-1327, 840-1045, 887-1605, 890-1054, 894-1126, 1010-1623, 1031-1302, 1052-1144, 1090-1723, 1140-2052, 1349-1637, 1372-1618, 1524-1761, 1594-2220, 1636-1965, 1659-1956, 1662-1946, 1748-2036, 1919-2156, 1919-2371, 1919-2397, 2062-2282, 2214-2912, 2256-2538, 2256-2569, 2285-2466, 2285-2593, 2397-2603, 2397-2605, 2424-2682, 2424-2698, 2427-2854, 2453-3065, 2462-2745, 2462-3197, 2495-2631, 2543-3211, 2784-3290, 2801-3060, 2924-3185, 2941-3188, 3072-3154, 3072-3221
13/4179651CB1/ 4279	1-695, 2-126, 7-281, 265-4279, 267-440, 301-523, 380-573, 389-1062, 551-1058, 560-1043, 560-1108, 857-1252, 938-1061, 994-1560, 1081-1379, 1169-1391, 1265-1556, 1328-1369, 1373-1993, 1393-1433, 1554-1786, 1554-2093, 1754-2218, 1756-2060, 1793-2052, 1825-2259, 1874-2127, 1875-2460, 1922-2189, 1932-2180, 1967-2451, 1969-2254, 1981-2640, 2042-2534, 2307-2541, 2357-2783, 2389-2670, 2451-2923, 2455-3049, 2464-2724, 2510-3099, 2562-3072, 2619-2856, 2662-3273, 2663-2858, 2666-3246, 2719-2946, 2747-2995, 2747-3269, 2778-3037, 2986-3292, 3139-3416, 3162-3390, 3162-3656, 3186-3469, 3187-3859, 3199-3501, 3210-3493, 3232-3800, 3280-3490, 3296-3818, 3333-3593, 3335-3617, 3340-3605, 3362-3609, 3362-3835, 3362-3843, 3406-3764, 3432-3860, 3525-3751, 3536-3781, 3536-3782, 3641-3905, 3695-3878, 3698-3866, 3698-3870, 3709-3953, 3709-4209, 3933-4202, 3953-4214, 3968-4107, 4033-4279

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
14/7482109CB1/ 1042	1-119, 70-257, 72-831, 677-1042, 678-1042, 680-1041, 680-1042
15/2867618CB1/ 4320	1-641, 1-653, 565-1278, 877-1420, 1055-1618, 1079-1693, 1106-1729, 1142-1922, 1188-1774, 1338-2100, 1358-1974, 1394-2074, 1478-1760, 1481-1878, 1760-2479, 1820-2098, 1841-2531, 1885-2045, 1977-2517, 1986-2165, 2001-2591, 2035-2507, 2057-2531, 2123-2532, 2140-2524, 2145-2747, 2151-2532, 2172-2532, 2221-2547, 2225-2493, 2330-3026, 2652-2874, 2706-3044, 2710-3396, 2856-3130, 2876-3159, 2892-3127, 2892-3451, 2918-3173, 3031-3596, 3070-3504, 3086-3338, 3108-3386, 3157-3493, 3174-3455, 3246-3549, 3255-3463, 3255-3707, 3308-3579, 3319-4013, 3321-4010, 3342-3695, 3526-4170, 3570-4148, 3575-4261, 3615-3869, 3617-3904, 3675-3942, 3701-3871, 3742-4032, 3742-4035, 3758-4296, 3776-4247, 3789-4264, 3814-4265, 3818-4268, 3820-4083, 3824-4102, 3834-4273, 3844-4046, 3844-4103, 3844-4234, 3846-4100, 3846-4320, 3847-4273, 3849-4267, 3858-4263, 3910-4277, 3938-4221, 3991-4222, 3991-4255, 3991-4266, 4002-4292, 4004-4260, 4037-4273, 4064-4272, 4075-4261, 4080-4268, 4085-4273, 4113-4264
16/7488348CB1/ 1733	1-1733, 33-1404
17/5539166CB1/ 2201	1-430, 209-939, 228-476, 239-1055, 262-859, 350-991, 453-1120, 457-1013, 457-1019, 611-1443, 682-869, 756-963, 756-1218, 756-1230, 820-1338, 1079-1530, 1411-1529, 1469-1998, 1469-2107, 1470-1758, 1470-1908, 1470-1943, 1470-2011, 1470-2065, 1470-2066, 1470-2201
18/7500341CB1/ 457	1-126, 1-218, 1-230, 1-457, 2-230, 12-126, 15-164, 16-230, 17-223, 19-230, 20-230, 22-168, 23-230, 28-175, 28-230, 29-230, 31-209, 36-445, 36-451, 37-202, 38-230, 42-202, 50-452, 228-428, 232-445, 310-446
19/7500846CB1/ 1513	1-268, 1-276, 1-495, 1-593, 1-1020, 4-266, 8-578, 13-221, 16-283, 112-643, 214-499, 246-678, 301-1011, 332-998, 384-648, 442-861, 654-803, 660-1023, 669-948, 669-972, 669-1013, 669-1023, 669-1032, 693-1023, 775-1020, 854-1513, 871-1111

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
20/3230921CB1/ 2084	1-240, 1-261, 1-266, 28-291, 28-650, 33-619, 36-332, 46-717, 47-896, 53-583, 115-354, 115-660, 160-297, 160-599, 160-818, 231-355, 231-369, 231-378, 231-422, 231-438, 238-438, 250-438, 254-438, 343-593, 439-568, 439-576, 439-582, 439-604, 439-613, 439-614, 439-619, 443-619, 444-572, 445-619, 449-619, 453-619, 456-619, 458-619, 466-619, 468-689, 468-795, 468-1004, 472-619, 483-1054, 483-1124, 497-616, 620-889, 620-1000, 620-1058, 620-1118, 641-1123, 646-1110, 653-892, 664-1126, 700-771, 716-1118, 723-1118, 750-1127, 751-1118, 759-1118, 931-1027, 957-1127, 1048-1118, 1068-1411, 1119-1915, 1120-1696, 1128-1677, 1144-1325, 1144-1446, 1144-1483, 1144-1605, 1144-1624, 1164-1408, 1164-1635, 1167-1652, 1200-1583, 1217-1333, 1223-1770, 1250-1893, 1261-1612, 1300-1691, 1304-1569, 1304-1750, 1319-1916, 1325-1920, 1401-1698, 1419-1722, 1419-1921, 1467-1727, 1482-1794, 1485-2084

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
11	7247177CB1	MIXDTXE01
12	5596327CB1	OVARNOT09
13	4179651CB1	BRAGNON02
15	2867618CB1	HIPONON01
17	5539166CB1	KIDNFEC01
18	7500341CB1	LNODNOT03
19	7500846CB1	PROSNOT06
20	3230921CB1	PANCNOT23

Table 6

Library	Vector	Library Description
BRAGNON02	pINCY	This normalized substantia nigra tissue library was constructed from 4.2 x 10^7 independent clones from a substantia nigra tissue library. Starting RNA was made from RNA isolated from substantia nigra tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus
		contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease. The library was normalized in two rounds using conditions adapted from Soares et al., 1994 (Proc. Natl. Acad. Sci. USA 91:9228-9232) and Bonaldo et al., 1996 (Genome Res. 6:791), except that a significantly longer (48 hours/round) reannealing hybridization was used.
HIPONON01	PSPORT1	This normalized hippocampus library was constructed from 1.13M independent clones from a hippocampus tissue library. RNA was isolated from the hippocampus tissue of a 72-year-old Caucasian female who died from an intracranial bleed. Patient history included nose cancer, hypertension, and arthritis. The normalization and hybridization conditions were adapted from Soares et al., 1994 (Proc. Natl. Acad. Sci. USA 91:9228-9232).
KIDNFEC01	PBLUESCRIPT	Library was constructed using RNA isolated from kidney tissue removed from a pool of twelve Caucasian male and female fetuses that were spontaneously aborted at 19-23 weeks' gestation.
LNODNOT03	pINCY	Library was constructed using RNA isolated from lymph node tissue obtained from a 67-year-old Caucasian male during a segmental lung resection and bronchoscopy. On microscopic exam, this tissue was found to be extensively necrotic with 10% viable tumor. Pathology for the associated tumor tissue indicated invasive grade 3-4 squamous cell carcinoma. Patient history included hemangioma. Family history included atherosclerotic coronary artery disease, benign hypertension, congestive heart failure, atherosclerotic coronary artery disease.

Table 6

Library	Vector	Library Description
MIXDTXE01	PBK-CMV	This 5' biased random primed library was constructed using pooled cDNA from nine donors. cDNA was generated using mRNA isolated from Jurkat cell line derived from the T cells of a male (donor A), THP-1 cell line derived from the peripheral blood of a 1-year-old male (donor B), Daudi cell line derived from B-lymphoblasts from a 16-year-old black male (donor C), RPMI-1666 cell line derived from lymphoma tissue from a 29-year-old Caucasian male (donor D), spleen from a 1-year-old Caucasian male (donor E), thymus removed from a 21-year-old Caucasian male (donor F) during a thymectomy, lymph node from a 42-year-old Caucasian female (donor G), thymus tumor from a 56-year-old Caucasian female (donor H) during a total thymectomy and PBMC's from a pool of donors (donor I). The patients presented with anemia and persistent hyperplastic thymus (H). Patient history included acute T-cell leukemia (A); acute monocytic leukemia (B); Burkitt's lymphoma (C); Hodgkin's disease (D); Bronchitis (E); hydrocele, regional enteritis or the small intestine, atopic dermatitis and benign neoplasm of the parathyroid (F); heart murmur and
		cardiac arrest (G); and cardiac dysrhythmia and left bundle branch block (H). Previous surgeries included an appendectomy and parathyroid surgery (F); unspecified heart surgery (G); and a normal delivery (H). Family history included benign hypertension in the grandparent(s) and coronary artery disease in the father of donor F.
OVARNOT09	pINCY	Library was constructed using RNA isolated from ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. Pathology indicated multiple follicular cysts ranging in size from 0.4 to 1.5 cm in the right and left ovaries, chronic cervicitis and squamous metaplasia of the cervix, and endometrium in weakly proliferative phase. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.
PANCNOT23	pINCY	Library was constructed using RNA isolated from diseased pancreatic tissue removed from a 43-year-old Caucasian female who died from a gunshot wound to the head. Patient history included type I diabetes and serology positive CMV antibody.
PROSNOT06	PSPORT1	Library was constructed using RNA isolated from the diseased prostate tissue of a 57-year-old Caucasian male during radical prostatectomy, removal of both testes and excision of regional lymph nodes. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated adenocarcinoma (Gleason grade 3+3). Patient history included a benign neoplasm of the large bowel and type I diabetes. Family history included a malignant neoplasm of the prostate and type I diabetes.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value=1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART, and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM, INCY, SMART, or TIGRFAM hits:* Probability value=1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-3, SEQ ID NO:5, and SEQ ID NO:7-10,

c) a polypeptide comprising a naturally occurring amino acid sequence at least 93% identical to the amino acid sequence of SEQ ID NO:4,

d) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to the amino acid sequence of SEQ ID NO:6,

e) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and

f) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

19. A method for treating a disease or condition associated with decreased expression of functional CHOP, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional CHOP, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23

and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional CHOP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of CHOP in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of CHOP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of CHOP

in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab

expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous

nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:11.

67. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:12.

68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID

NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

<110> INCYTE GENOMICS, INC.
SWARNAKAR, Anita
GORVAD, Ann E.
HAFALIA, April J.A.
DUGGAN, Brendan M.
EMERLING, Brooke M.
ISON, Craig H.
NGUYEN, Danniel B.
LEE, Ernestine A.
YUE, Henry
YUE, Huibin
FORSYTHE, Ian J.
LI, Joana X.
THANGAVELU,Kavitha
WALIA, Narinder K.
BURFORD, Neil
MASON, Patricia M.
LAL, Preeti G.
LEE, Sally
BECHA, Shanya D.
TANG, Y. Tom

<120> Carbohydrate-Associated Proteins

<130> PF-0982 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/293,768; 60/309,548; 60/314,400; 60/343,706; 60/337,999
<151> 2001-05-25; 2001-08-01; 2001-08-23; 2001-10-19; 2001-12-07

<160> 20

<170> PERL Program

<210> 1
<211> 912
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7247177CD1

<400> 1
Met Glu Ala Ala Val Lys Glu Glu Ile Ser Val Glu Asp Glu Ala
1 5 10 15
Val Asp Lys Asn Ile Phe Arg Asp Cys Asn Lys Ile Ala Phe Tyr
20 25 30
Arg Arg Gln Lys Gln Trp Leu Ser Lys Lys Ser Thr Tyr Arg Ala
35 40 45
Leu Leu Asp Ser Val Thr Thr Asp Glu Asp Ser Thr Arg Phe Gln
50 55 60
Ile Ile Asn Glu Ala Ser Lys Val Pro Leu Leu Ala Glu Ile Tyr
65 70 75
Gly Ile Glu Gly Asn Ile Phe Arg Leu Lys Ile Asn Glu Glu Thr
80 85 90
Pro Leu Lys Pro Arg Phe Glu Val Pro Asp Val Leu Thr Ser Lys
95 100 105
Pro Ser Thr Val Arg Leu Ile Ser Cys Ser Gly Asp Thr Gly Ser
110 115 120
Leu Ile Leu Ala Asp Gly Lys Gly Asp Leu Lys Cys His Ile Thr

```
                125                 130                 135
Ala Asn Pro Phe Lys Val Asp Leu Val Ser Glu Glu Glu Val Val
            140                 145                 150
Ile Ser Ile Asn Ser Leu Gly Gln Leu Tyr Phe Glu His Leu Gln
            155                 160                 165
Ile Leu His Lys Gln Arg Ala Ala Lys Glu Asn Glu Glu Glu Thr
            170                 175                 180
Ser Val Asp Thr Ser Gln Glu Asn Gln Glu Asp Leu Gly Leu Trp
            185                 190                 195
Glu Glu Lys Phe Gly Lys Phe Val Asp Ile Lys Ala Asn Gly Pro
            200                 205                 210
Ser Ser Ile Gly Leu Asp Phe Ser Leu His Gly Phe Glu His Leu
            215                 220                 225
Tyr Gly Ile Pro Gln His Ala Glu Ser His Gln Leu Lys Asn Thr
            230                 235                 240
Gly Asp Gly Asp Ala Tyr Arg Leu Tyr Asn Leu Asp Val Tyr Gly
            245                 250                 255
Tyr Gln Ile Tyr Asp Lys Met Gly Ile Tyr Gly Ser Val Pro Tyr
            260                 265                 270
Leu Leu Ala His Lys Leu Gly Arg Thr Ile Gly Ile Phe Trp Leu
            275                 280                 285
Asn Ala Ser Glu Thr Leu Val Glu Ile Asn Thr Glu Pro Ala Val
            290                 295                 300
Glu Tyr Thr Leu Thr Gln Met Gly Pro Val Ala Ala Lys Gln Lys
            305                 310                 315
Val Arg Ser Arg Thr His Val His Trp Met Ser Glu Ser Gly Ile
            320                 325                 330
Ile Asp Val Phe Leu Leu Thr Gly Pro Thr Pro Ser Asp Val Phe
            335                 340                 345
Lys Gln Tyr Ser His Leu Thr Gly Thr Gln Ala Met Pro Pro Leu
            350                 355                 360
Phe Ser Leu Gly Tyr His Gln Cys Arg Trp Asn Tyr Glu Asp Glu
            365                 370                 375
Gln Asp Val Lys Ala Val Asp Ala Gly Phe Asp Glu His Asp Ile
            380                 385                 390
Pro Tyr Asp Ala Met Trp Leu Asp Ile Glu His Thr Glu Gly Lys
            395                 400                 405
Arg Tyr Phe Thr Trp Asp Lys Asn Arg Phe Pro Asn Pro Lys Arg
            410                 415                 420
Met Gln Glu Leu Leu Arg Ser Lys Lys Arg Lys Leu Val Val Ile
            425                 430                 435
Ser Asp Pro His Ile Lys Ile Asp Pro Asp Tyr Ser Val Tyr Val
            440                 445                 450
Lys Ala Lys Asp Gln Gly Phe Phe Val Lys Asn Gln Glu Gly Glu
            455                 460                 465
Asp Phe Glu Gly Val Cys Trp Pro Gly Leu Ser Ser Tyr Leu Asp
            470                 475                 480
Phe Thr Asn Pro Lys Val Arg Glu Trp Tyr Ser Ser Leu Phe Ala
            485                 490                 495
Phe Pro Val Tyr Gln Gly Ser Thr Asp Ile Leu Phe Leu Trp Asn
            500                 505                 510
Asp Met Asn Glu Pro Ser Val Phe Arg Gly Pro Glu Gln Thr Met
            515                 520                 525
Gln Lys Asn Ala Ile His His Gly Asn Trp Glu His Arg Glu Leu
            530                 535                 540
His Asn Ile Tyr Gly Phe Tyr His Gln Met Ala Thr Ala Glu Gly
            545                 550                 555
Leu Ile Lys Arg Ser Lys Gly Lys Glu Arg Pro Phe Val Leu Thr
            560                 565                 570
Arg Ser Phe Phe Ala Gly Ser Gln Lys Tyr Gly Ala Val Trp Thr
            575                 580                 585
Gly Asp Asn Thr Ala Glu Trp Ser Asn Leu Lys Ile Ser Ile Pro
            590                 595                 600
```

```
Met Leu Leu Thr Leu Ser Ile Thr Gly Ile Ser Phe Cys Gly Ala
                605                 610                 615
Asp Ile Gly Gly Phe Ile Gly Asn Pro Glu Thr Glu Leu Leu Val
                620                 625                 630
Arg Trp Tyr Gln Ala Gly Ala Tyr Gln Pro Phe Phe Arg Gly His
                635                 640                 645
Ala Thr Met Asn Thr Lys Arg Arg Ala Pro Trp Leu Phe Gly Glu
                650                 655                 660
Glu His Thr Arg Leu Ile Arg Glu Ala Ile Arg Glu Arg Tyr Gly
                665                 670                 675
Leu Leu Pro Tyr Trp Tyr Ser Leu Phe Tyr His Ala His Val Ala
                680                 685                 690
Ser Gln Pro Val Met Arg Pro Leu Trp Val Glu Phe Pro Asp Glu
                695                 700                 705
Leu Lys Thr Phe Asp Met Glu Asp Glu Tyr Met Leu Gly Ser Ala
                710                 715                 720
Leu Leu Val His Pro Val Thr Glu Pro Lys Ala Thr Thr Val Asp
                725                 730                 735
Val Phe Leu Pro Gly Ser Asn Glu Val Trp Tyr Asp Tyr Lys Thr
                740                 745                 750
Phe Ala His Trp Glu Gly Gly Cys Thr Val Lys Ile Pro Val Ala
                755                 760                 765
Leu Asp Thr Ile Pro Val Phe Gln Arg Gly Gly Ser Val Ile Pro
                770                 775                 780
Ile Lys Thr Thr Val Gly Lys Ser Thr Gly Trp Met Thr Glu Ser
                785                 790                 795
Ser Tyr Gly Leu Arg Val Ala Leu Ser Thr Lys Gly Ser Ser Val
                800                 805                 810
Gly Glu Leu Tyr Leu Asp Asp Gly His Ser Phe Gln Tyr Leu His
                815                 820                 825
Gln Lys Gln Phe Leu His Arg Lys Phe Ser Phe Cys Ser Ser Val
                830                 835                 840
Leu Ile Asn Ser Ser Ala Asp Gln Arg Gly His Tyr Pro Ser Lys
                845                 850                 855
Cys Val Val Glu Lys Ile Leu Val Leu Gly Phe Arg Lys Glu Pro
                860                 865                 870
Ser Ser Val Thr Thr His Ser Ser Asp Gly Lys Asp Gln Pro Val
                875                 880                 885
Ala Phe Thr Tyr Cys Ala Lys Thr Ser Ile Leu Ser Leu Glu Lys
                890                 895                 900
Leu Ser Leu Asn Ile Ala Thr Asp Trp Glu Gly Ser
                905                 910
```

<210> 2
<211> 691
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5596327CD1

<400> 2

```
Met Leu Gln Glu Glu Pro Asp Leu Val Ser Ala Ile Tyr Gly Arg
  1               5                  10                  15
Gly Ile Ala Tyr Gly Lys Lys Gly Leu His Asp Ile Lys Asn Ala
                20                  25                  30
Glu Leu Ala Leu Phe Glu Leu Ser Arg Val Ile Thr Leu Glu Pro
                35                  40                  45
Asp Arg Pro Glu Val Phe Glu Gln Arg Ala Glu Ile Leu Ser Pro
                50                  55                  60
Leu Gly Arg Ile Asn Glu Ala Val Asn Asp Leu Thr Lys Ala Ile
                65                  70                  75
```

```
Gln Leu Gln Pro Ser Ala Arg Leu Tyr Arg His Arg Gly Thr Leu
                80                  85                  90
Tyr Phe Ile Ser Glu Asp Tyr Ala Thr Ala His Glu Asp Phe Gln
            95                  100                 105
Gln Ser Leu Glu Leu Asn Lys Asn Gln Pro Ile Ala Met Leu Tyr
        110                 115                 120
Lys Gly Leu Thr Phe Phe His Arg Gly Leu Leu Lys Glu Ala Ile
    125                 130                 135
Glu Ser Phe Lys Glu Ala Leu Lys Gln Lys Val Asp Phe Ile Asp
140                 145                 150
Ala Tyr Lys Ser Leu Gly Gln Ala Tyr Arg Glu Leu Gly Asn Phe
155                 160                 165
Glu Ala Ala Thr Glu Ser Phe Gln Lys Ala Leu Leu Leu Asn Gln
170                 175                 180
Asn His Val Gln Thr Leu Gln Leu Arg Gly Met Met Leu Tyr His
185                 190                 195
His Gly Ser Leu Gln Glu Ala Leu Lys Asn Phe Lys Arg Cys Leu
200                 205                 210
Gln Leu Glu Pro Tyr Asn Glu Val Cys Gln Tyr Met Lys Gly Leu
215                 220                 225
Ser His Val Ala Met Gly Gln Phe Tyr Glu Gly Ile Lys Ala Gln
230                 235                 240
Thr Lys Val Met Leu Asn Asp Pro Leu Pro Gly Gln Lys Ala Ser
245                 250                 255
Pro Glu Tyr Leu Lys Val Lys Tyr Leu Arg Glu Tyr Ser Arg Tyr
260                 265                 270
Leu His Ala His Leu Asp Thr Pro Leu Thr Glu Tyr Asn Ile Asp
275                 280                 285
Val Asp Leu Pro Gly Ser Phe Lys Asp His Trp Ala Lys Asn Leu
290                 295                 300
Pro Phe Leu Ile Glu Asp Tyr Glu Glu Gln Pro Gly Leu Gln Pro
305                 310                 315
His Ile Lys Asp Val Leu His Gln Asn Phe Glu Ser Tyr Lys Pro
320                 325                 330
Glu Val Gln Glu Leu Ile Cys Val Ala Asp Arg Leu Gly Ser Leu
335                 340                 345
Met Gln Tyr Glu Thr Pro Gly Phe Leu Pro Asn Lys Arg Ile His
350                 355                 360
Arg Ala Met Gly Leu Ala Ala Leu Glu Val Met Gln Ala Val Gln
365                 370                 375
Arg Thr Trp Thr Asn Ser Lys Val Arg Met Asn Gly Lys Thr Arg
380                 385                 390
Leu Met Gln Trp Arg Asp Met Phe Asp Ile Ala Val Lys Trp Arg
395                 400                 405
Arg Ile Ala Asp Pro Asp Gln Pro Val Leu Trp Leu Asp Gln Met
410                 415                 420
Pro Ala Arg Ser Leu Ser Arg Gly Phe Asn Asn His Ile Asn Leu
425                 430                 435
Ile Arg Gly Gln Val Ile Asn Met Arg Tyr Leu Glu Tyr Phe Glu
440                 445                 450
Lys Ile Leu His Phe Ile Lys Asp Arg Ile Leu Val Tyr His Gly
455                 460                 465
Ala Asn Asn Pro Lys Gly Leu Leu Glu Val Arg Glu Ala Leu Glu
470                 475                 480
Lys Val His Lys Val Glu Asp Leu Leu Pro Ile Met Lys Phe Asn
485                 490                 495
Thr Lys Thr Lys Asp Gly Phe Thr Val Asn Thr Lys Val Pro Ser
500                 505                 510
Leu Lys Asp Gln Gly Lys Glu Tyr Asp Gly Phe Thr Ile Thr Ile
515                 520                 525
Thr Gly Asp Lys Val Gly Asn Ile Leu Phe Ser Val Glu Thr Gln
530                 535                 540
Thr Thr Glu Glu Arg Thr Gln Leu Tyr His Ala Glu Ile Asp Ala
```

```
                545                 550                 555
Leu Tyr Lys Asp Leu Thr Ala Lys Gly Lys Val Leu Ile Leu Ser
                560                 565                 570
Ser Glu Phe Gly Glu Ala Asp Ala Val Cys Asn Leu Ile Leu Ser
                575                 580                 585
Leu Val Tyr Tyr Phe Tyr Asn Leu Met Pro Leu Ser Arg Gly Ser
                590                 595                 600
Ser Val Ile Ala Tyr Ser Val Ile Val Gly Ala Leu Met Ala Ser
                605                 610                 615
Gly Lys Glu Val Ala Gly Lys Ile Pro Lys Gly Lys Leu Val Asp
                620                 625                 630
Phe Glu Ala Met Thr Ala Pro Gly Ser Glu Ala Phe Ser Lys Val
                635                 640                 645
Ala Lys Ser Trp Met Asn Leu Lys Ser Ile Ser Pro Ser Tyr Lys
                650                 655                 660
Thr Leu Pro Ser Val Ser Glu Thr Phe Pro Thr Leu Arg Ser Met
                665                 670                 675
Ile Glu Val Leu Asn Thr Asp Ser Ser Pro Arg Cys Leu Lys Lys
                680                 685                 690
Leu
```

<210> 3
<211> 506
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4179651CD1

<400> 3

```
Met Thr Asp Ala Glu Arg Val Asp Gln Ala Tyr Arg Glu Asn Gly
  1                 5                  10                  15
Phe Asn Ile Tyr Val Ser Asp Lys Ile Ser Leu Asn Arg Ser Leu
                 20                  25                  30
Pro Asp Ile Arg His Pro Asn Cys Asn Ser Lys Arg Tyr Leu Glu
                 35                  40                  45
Thr Leu Pro Asn Thr Ser Ile Ile Ile Pro Phe His Asn Glu Gly
                 50                  55                  60
Trp Ser Ser Leu Leu Arg Thr Val His Ser Val Leu Asn Arg Ser
                 65                  70                  75
Pro Pro Glu Leu Val Ala Glu Ile Val Leu Val Asp Asp Phe Ser
                 80                  85                  90
Asp Arg Glu His Leu Lys Lys Pro Leu Glu Asp Tyr Met Ala Leu
                 95                 100                 105
Phe Pro Ser Val Arg Ile Leu Arg Thr Lys Lys Arg Glu Gly Leu
                110                 115                 120
Ile Arg Thr Arg Met Leu Gly Ala Ser Val Ala Thr Gly Asp Val
                125                 130                 135
Ile Thr Phe Leu Asp Ser His Cys Glu Ala Asn Val Asn Trp Leu
                140                 145                 150
Pro Pro Leu Leu Asp Arg Ile Ala Arg Asn Arg Lys Thr Ile Val
                155                 160                 165
Cys Pro Met Ile Asp Val Ile Asp His Asp Asp Phe Arg Tyr Glu
                170                 175                 180
Thr Gln Ala Gly Asp Ala Met Arg Gly Ala Phe Asp Trp Glu Met
                185                 190                 195
Tyr Tyr Lys Arg Ile Pro Ile Pro Pro Glu Leu Gln Lys Ala Asp
                200                 205                 210
Pro Ser Asp Pro Phe Glu Ser Pro Val Met Ala Gly Gly Leu Phe
                215                 220                 225
Ala Val Asp Arg Lys Trp Phe Trp Glu Leu Gly Gly Tyr Asp Pro
```

```
                230                 235                 240
Gly Leu Glu Ile Trp Gly Gly Glu Gln Tyr Glu Ile Ser Phe Lys
                245                 250                 255
Val Trp Met Cys Gly Gly Arg Met Glu Asp Ile Pro Cys Ser Arg
                260                 265                 270
Val Gly His Ile Tyr Arg Lys Tyr Val Pro Tyr Lys Val Pro Ala
                275                 280                 285
Gly Val Ser Leu Ala Arg Asn Leu Lys Arg Val Ala Glu Val Trp
                290                 295                 300
Met Asp Glu Tyr Ala Glu Tyr Ile Tyr Gln Arg Arg Pro Glu Tyr
                305                 310                 315
Arg His Leu Ser Ala Gly Asp Val Ala Val Gln Lys Lys Leu Arg
                320                 325                 330
Ser Ser Leu Asn Cys Lys Ser Phe Lys Trp Phe Met Thr Lys Ile
                335                 340                 345
Ala Trp Asp Leu Pro Lys Phe Tyr Pro Pro Val Glu Pro Pro Ala
                350                 355                 360
Ala Ala Trp Gly Glu Ile Arg Asn Val Gly Thr Gly Leu Cys Ala
                365                 370                 375
Asp Thr Lys His Gly Ala Leu Gly Ser Pro Leu Arg Leu Glu Gly
                380                 385                 390
Cys Val Arg Gly Arg Gly Glu Ala Ala Trp Asn Asn Met Gln Val
                395                 400                 405
Phe Thr Phe Thr Trp Arg Glu Asp Ile Arg Pro Gly Asp Pro Gln
                410                 415                 420
His Thr Lys Lys Phe Cys Phe Asp Ala Ile Ser His Thr Ser Pro
                425                 430                 435
Val Thr Leu Tyr Asp Cys His Ser Met Lys Gly Asn Gln Leu Trp
                440                 445                 450
Lys Tyr Arg Lys Asp Lys Thr Leu Tyr His Pro Val Ser Gly Ser
                455                 460                 465
Cys Met Asp Cys Ser Glu Ser Asp His Arg Ile Phe Met Asn Thr
                470                 475                 480
Cys Asn Pro Ser Ser Leu Thr Gln Gln Trp Leu Phe Glu His Thr
                485                 490                 495
Asn Ser Thr Val Leu Glu Lys Phe Asn Arg Asn
                500                 505

<210> 4
<211> 276
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7482109CD1

<400> 4
Met Asn Ser Lys Arg Met Leu Leu Leu Val Leu Phe Ala Phe Ser
  1                 5                  10                  15
Leu Met Leu Val Glu Arg Tyr Phe Arg Asn His Gln Val Glu Glu
                 20                  25                  30
Leu Arg Leu Ser Asp Trp Phe His Pro Arg Lys Arg Pro Asp Val
                 35                  40                  45
Ile Thr Lys Thr Asp Trp Leu Ala Pro Val Leu Trp Glu Gly Thr
                 50                  55                  60
Phe Asp Arg Arg Val Leu Glu Lys His Tyr Arg Arg Arg Asn Ile
                 65                  70                  75
Thr Val Gly Leu Ala Val Phe Ala Thr Gly Arg Phe Ala Glu Glu
                 80                  85                  90
Tyr Leu Arg Pro Phe Leu His Ser Ala Asn Lys His Phe Met Thr
                 95                 100                 105
Gly Tyr Arg Val Ile Phe Tyr Ile Met Val Asp Ala Phe Phe Lys
```

```
                110                 115                 120
Leu Pro Asp Ile Glu Pro Ser Pro Leu Arg Thr Phe Lys Ala Phe
                125                 130                 135
Lys Val Gly Thr Glu Arg Trp Trp Leu Asp Gly Pro Leu Val His
                140                 145                 150
Val Lys Ser Leu Gly Glu His Ile Ala Ser His Ile Gln Asp Glu
                155                 160                 165
Val Asp Phe Leu Phe Ser Met Ala Ala Asn Gln Val Phe Gln Asn
                170                 175                 180
Glu Phe Gly Val Glu Thr Leu Gly Pro Leu Val Ala Gln Leu His
                185                 190                 195
Ala Trp Trp Tyr Phe Arg Asn Thr Lys Asn Phe Pro Tyr Glu Arg
                200                 205                 210
Arg Pro Thr Ser Ala Ala Cys Ile Pro Phe Gly Gln Gly Asp Phe
                215                 220                 225
Tyr Tyr Gly Asn Leu Met Val Gly Gly Thr Pro His Asn Ile Leu
                230                 235                 240
Asp Phe Ile Lys Glu Tyr Leu Asn Gly Val Ile His Asp Ile Lys
                245                 250                 255
Asn Gly Leu Asn Ser Thr Tyr Glu Lys His Leu Asn Lys Tyr Phe
                260                 265                 270
Tyr Leu Asn Lys Pro Thr
                275
```

<210> 5
<211> 378
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2867618CD1

<400> 5

```
Met Lys Pro Ser Lys Ala Ser Glu Leu Asn Leu Asp Glu Leu Pro
  1                 5                  10                  15
Pro Leu Asn Ile Tyr Leu His Val Phe Tyr Tyr Ser Trp Tyr Gly
                 20                  25                  30
Asn Pro Gln Phe Asp Gly Lys Tyr Ile His Trp Asn His Pro Val
                 35                  40                  45
Leu Glu His Trp Asp Pro Arg Ile Ala Lys Asn Tyr Pro Gln Gly
                 50                  55                  60
Arg His Asn Pro Pro Asp Asp Ile Gly Ser Ser Phe Tyr Pro Glu
                 65                  70                  75
Leu Gly Ser Tyr Ser Ser Arg Asp Pro Ser Val Ile Glu Thr His
                 80                  85                  90
Met Arg Gln Met Arg Ser Ala Ser Ile Gly Val Leu Ala Leu Ser
                 95                 100                 105
Trp Tyr Pro Pro Asp Val Asn Asp Glu Asn Gly Glu Pro Thr Asp
                110                 115                 120
Asn Leu Val Pro Thr Ile Leu Asp Lys Ala His Lys Tyr Asn Leu
                125                 130                 135
Lys Val Thr Phe His Ile Glu Pro Tyr Ser Ile Glu Met Ile Lys
                140                 145                 150
His Val Gln Asn Val Lys Tyr Ile Ile Asp Lys Tyr Gly Asn His
                155                 160                 165
Pro Ala Phe Tyr Arg Tyr Lys Thr Lys Thr Gly Asn Ala Leu Pro
                170                 175                 180
Met Phe Tyr Val Tyr Asp Ser Tyr Val Thr Lys Pro Glu Lys Arg
                185                 190                 195
Ala Asn Leu Leu Thr Thr Ser Gly Ser Arg Ser Ile Arg Asn Ser
                200                 205                 210
Pro Tyr Gly Gly Leu Phe Ile Ala Leu Leu Val Glu Glu Glu His
```

```
                215                 220                 225
Lys Tyr Asp Ile Leu Gln Arg Val Asp Gly Ile Tyr Thr Tyr Phe
                230                 235                 240
Ala Thr Asn Gly Phe Thr Tyr Gly Ser Ser His Gln Asn Gly Ala
                245                 250                 255
Ser Leu Lys Leu Leu Cys Asp Lys Tyr Asn Leu Ile Phe Ile Pro
                260                 265                 270
Ser Val Gly Pro Gly Tyr Ile Asp Thr Ser Ile Arg Pro Trp Asn
                275                 280                 285
Thr Gln Asn Thr Arg Asn Arg Ile Asn Gly Lys Tyr Tyr Glu Ile
                290                 295                 300
Gly Leu Ser Ala Ala Leu Gln Thr Arg Pro Ser Leu Ile Ser Ile
                305                 310                 315
Thr Ser Phe Asn Glu Trp His Glu Gly Thr Gln Ile Glu Lys Ala
                320                 325                 330
Val Pro Lys Arg Thr Ser Asn Thr Val Tyr Leu Asp Tyr Arg Pro
                335                 340                 345
His Lys Pro Gly Leu Tyr Leu Glu Leu Thr Arg Lys Trp Ser Glu
                350                 355                 360
Lys Tyr Ser Lys Glu Arg Ala Thr Tyr Ala Leu Asp Arg Gln Leu
                365                 370                 375
Pro Val Ser
```

<210> 6
<211> 458
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7488348CD1

<400> 6

```
Met Ser Ile Leu Lys Ile Ile His Ala Arg Asp Ile Phe Glu Ser
  1                 5                  10                  15
Arg Gly Asn Pro Thr Val Glu Val Asp Leu Tyr Thr Asn Lys Gly
                 20                  25                  30
Gly Leu Phe Gly Arg Ala Ala Val Pro Ser Gly Ala Ser Thr Gly
                 35                  40                  45
Ile Tyr Glu Ala Leu Leu Glu Leu Arg Asp Asn Asp Lys Thr Arg
                 50                  55                  60
Tyr Met Gly Gly Lys Gly Val Ser Lys Ala Val Glu His Ile Ile
                 65                  70                  75
Asn Lys Thr Ile Ala Pro Ala Leu Ile Ser Lys Asn Val Asn Val
                 80                  85                  90
Val Glu Gln Asp Lys Ile Asp Asn Leu Met Leu Asp Met Asp Gly
                 95                 100                 105
Ser Glu Asn Lys Ser Lys Phe Gly Ala Asn Ala Ile Leu Gly Val
                110                 115                 120
Ser Leu Ala Val Cys Ser Asn Ala Gly Ala Thr Ala Glu Lys Gly
                125                 130                 135
Val Pro Leu Tyr Arg His Ile Ala Asp Leu Ala Gly Asn Asn Pro
                140                 145                 150
Glu Val Ile Leu Pro Val Pro Ala Phe Asn Val Ile Asn Gly Gly
                155                 160                 165
Ser His Ala Gly Asn Lys Leu Ala Met Gln Glu Phe Met Ile Pro
                170                 175                 180
Pro Cys Gly Ala Asp Arg Phe Asn Asp Ala Ile Arg Ile Gly Ala
                185                 190                 195
Glu Val Tyr His Asn Leu Lys Asn Val Ile Lys Glu Lys Tyr Gly
                200                 205                 210
Lys Asp Ala Thr Asn Val Gly Asp Glu Gly Gly Phe Ala Pro Asn
```

```
                215                 220                 225
Ile Leu Glu Asn Lys Glu Gly Leu Glu Leu Leu Lys Thr Ala Ile
                230                 235                 240
Gly Lys Ala Gly Tyr Thr Asp Lys Val Val Ile Gly Met Asp Val
                245                 250                 255
Ala Ala Ser Glu Phe Phe Arg Ser Gly Lys Tyr Asp Leu Asp Phe
                260                 265                 270
Lys Ser Pro Asp Asp Pro Ser Arg Tyr Ile Ser Pro Asp Gln Leu
                275                 280                 285
Ala Asp Leu Tyr Lys Gly Phe Val Leu Gly His Ala Val Lys Asn
                290                 295                 300
Tyr Pro Val Gly Val Ser Ile Glu Asp Pro Pro Phe Asp Gln Asp
                305                 310                 315
Asp Trp Gly Ala Trp Lys Lys Leu Phe Thr Gly Ser Leu Val Gly
                320                 325                 330
Ile Gln Val Val Gly Asp Asp Leu Thr Val Thr Lys Pro Glu Ala
                335                 340                 345
Arg Ile Ala Lys Ala Val Glu Glu Val Lys Ala Cys Asn Cys Leu
                350                 355                 360
Leu Leu Leu Lys Val Asn Gln Ile Gly Ser Val Thr Glu Ser Leu
                365                 370                 375
Gln Ala Cys Lys Leu Ala Gln Ser Asn Gly Trp Gly Val Met Pro
                380                 385                 390
Val Ser His Arg Leu Ser Gly Glu Thr Glu Asp Thr Phe Met Ala
                395                 400                 405
Asp Leu Val Val Gly Leu Cys Thr Gly Gln Ile Lys Thr Gly Pro
                410                 415                 420
Thr Cys Arg Ser Glu Arg Leu Ala Lys Tyr Asn Gln Leu Leu Arg
                425                 430                 435
Ile Glu Glu Ala Glu Ala Gly Ser Lys Ala Arg Phe Ala Gly Arg
                440                 445                 450
Asn Phe Arg Asn Pro Arg Ile Asn
                455

<210> 7
<211> 402
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5539166CD1

<400> 7
Met Met Gly Ser Trp Lys His Cys Leu Phe Ser Ala Ser Leu Ile
  1                 5                  10                  15
Ser Ala Leu Ile Phe Val Phe Val Tyr Asn Thr Glu Leu Trp Glu
                 20                  25                  30
Asn Lys Arg Phe Leu Arg Ala Ala Leu Ser Asn Ala Ser Leu Leu
                 35                  40                  45
Ala Glu Ala Cys His Gln Ile Phe Glu Gly Lys Val Phe Tyr Pro
                 50                  55                  60
Thr Glu Asn Ala Leu Lys Thr Thr Leu Asp Glu Ala Thr Cys Tyr
                 65                  70                  75
Glu Tyr Met Val Arg Ser His Tyr Val Thr Glu Thr Leu Ser Glu
                 80                  85                  90
Glu Glu Ala Gly Phe Pro Leu Ala Tyr Thr Val Thr Ile His Lys
                 95                 100                 105
Asp Phe Gly Thr Phe Glu Arg Leu Phe Arg Ala Ile Tyr Met Pro
                110                 115                 120
Gln Asn Val Tyr Cys Val His Leu Asp Gln Lys Ala Thr Asp Ala
                125                 130                 135
Phe Lys Gly Ala Val Lys Gln Leu Leu Ser Cys Phe Pro Asn Ala
```

```
                 140                 145                 150
Phe Leu Ala Ser Lys Lys Glu Ser Val Val Tyr Gly Gly Ile Ser
                 155                 160                 165
Arg Leu Gln Ala Asp Leu Asn Cys Leu Glu Asp Leu Val Ala Ser
                 170                 175                 180
Glu Val Pro Trp Lys Tyr Val Ile Asn Thr Cys Gly Gln Asp Phe
                 185                 190                 195
Pro Leu Lys Thr Asn Arg Glu Ile Val Gln Tyr Leu Lys Gly Phe
                 200                 205                 210
Lys Gly Lys Asn Ile Thr Pro Gly Val Leu Pro Pro Asp His Ala
                 215                 220                 225
Val Gly Arg Thr Lys Tyr Val His Gln Glu Leu Leu Asn His Lys
                 230                 235                 240
Asn Ser Tyr Val Ile Lys Thr Thr Lys Leu Lys Thr Pro Pro Pro
                 245                 250                 255
His Asp Met Val Ile Tyr Leu Gly Thr Ala Tyr Val Ala Leu Thr
                 260                 265                 270
Arg Asp Phe Ala Asn Phe Val Leu Gln Asp Gln Leu Ala Leu Asp
                 275                 280                 285
Leu Leu Ser Trp Ser Lys Asp Thr Tyr Ser Pro Asp Glu His Phe
                 290                 295                 300
Trp Val Thr Leu Asn Arg Ile Pro Gly Val Pro Gly Ser Met Pro
                 305                 310                 315
Asn Ala Ser Trp Thr Gly Asn Leu Arg Ala Ile Lys Trp Ser Asp
                 320                 325                 330
Met Glu Asp Arg His Gly Gly Cys His Gly His Tyr Val His Gly
                 335                 340                 345
Ile Cys Ile Tyr Gly Asn Gly Asp Leu Lys Trp Leu Val Asn Ser
                 350                 355                 360
Pro Ser Leu Phe Ala Asn Lys Phe Glu Leu Asn Thr Tyr Pro Leu
                 365                 370                 375
Thr Val Glu Cys Leu Glu Leu Arg His Arg Glu Arg Thr Leu Asn
                 380                 385                 390
Gln Ser Glu Thr Ala Ile Gln Pro Ser Trp Tyr Phe
                 395                 400
```

<210> 8
<211> 109
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500341CD1

<400> 8

```
Met Ala Thr Gly Thr Asp Gln Val Val Gly Leu Gly Leu Val Ala
  1               5                  10                  15
Val Ser Leu Ile Ile Phe Thr Tyr Tyr Thr Ala Trp Val Ile Leu
                 20                  25                  30
Leu Pro Phe Ile Asp Ser Gln His Val Ile His Lys Tyr Phe Leu
                 35                  40                  45
Pro Arg Ala Tyr Ala Val Ala Ile Pro Leu Ala Ala Gly Leu Leu
                 50                  55                  60
Leu Leu Leu Phe Val Gly Ser Phe Pro Tyr Pro Asp Thr Ala Pro
                 65                  70                  75
Asp Pro Thr Lys Pro Ser Gly Pro Gly Lys Pro Gly Glu Gly Leu
                 80                  85                  90
Thr Asp Pro Arg Thr Ser Pro Gly Ala Gln Gly Ser His Pro Gly
                 95                 100                 105
Pro Leu Thr Asp
```

<210> 9
<211> 225
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500846CD1

<400> 9
Met Leu Ser Met Leu Arg Thr Met Thr Arg Leu Cys Phe Leu Leu
1 5 10 15
Phe Phe Ser Val Ala Thr Ser Gly Cys Ser Ala Ala Ala Ala Ser
20 25 30
Ser Leu Glu Met Leu Ser Arg Glu Phe Glu Thr Cys Ala Phe Ser
35 40 45
Phe Ser Ser Leu Pro Arg Ser Cys Lys Glu Ile Lys Glu Arg Cys
50 55 60
His Ser Ala Gly Asp Gly Leu Tyr Phe Leu Arg Thr Lys Asn Gly
65 70 75
Val Val Tyr Gln Thr Phe Cys Asp Met Thr Ser Gly Gly Gly Gly
80 85 90
Trp Thr Leu Val Ala Ser Val His Glu Asn Asp Met Arg Gly Lys
95 100 105
Cys Thr Val Gly Asp Arg Trp Ser Ser Gln Gln Gly Asn Lys Ala
110 115 120
Asp Tyr Pro Glu Gly Asp Gly Asn Trp Ala Asn Tyr Asn Thr Phe
125 130 135
Gly Ser Ala Glu Ala Ala Thr Ser Asp Asp Tyr Lys Asn Pro Gly
140 145 150
Tyr Tyr Asp Ile Gln Ala Lys Asp Leu Gly Ile Trp His Val Pro
155 160 165
Asn Lys Ser Pro Met Gln His Trp Arg Asn Ser Ala Leu Leu Arg
170 175 180
Tyr Arg Thr Asn Thr Gly Phe Leu Gln Arg Leu Gly His Asn Leu
185 190 195
Phe Gly Ile Tyr Gln Gly Asn Leu Leu Gln Asp Ser Phe Ser Ser
200 205 210
Gly Cys Leu Ile Thr Arg Glu Gln Pro Thr Pro Phe Val Leu Gly
215 220 225

<210> 10
<211> 463
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3230921CD1

<400> 10
Met Ser Glu Glu Glu Ala Ala Gln Ile Pro Arg Ser Ser Val Trp
1 5 10 15
Glu Gln Asp Gln Gln Asn Val Val Gln Arg Val Val Ala Leu Pro
20 25 30
Leu Val Arg Ala Thr Cys Thr Ala Val Cys Asp Val Tyr Ser Ala
35 40 45
Ala Lys Asp Arg His Pro Leu Leu Gly Ser Ala Cys Arg Leu Ala
50 55 60
Glu Asn Cys Val Cys Gly Leu Thr Thr Arg Ala Leu Asp His Ala
65 70 75
Gln Pro Leu Leu Glu His Leu Gln Pro Gln Leu Ala Thr Met Asn

```
                80                  85                  90
Ser Leu Ala Cys Arg Gly Leu Asp Lys Leu Glu Glu Lys Leu Pro
                95                 100                 105
Phe Leu Gln Gln Pro Ser Glu Thr Val Val Thr Ser Ala Lys Asp
               110                 115                 120
Val Val Ala Ser Ser Val Thr Gly Val Val Asp Leu Ala Arg Arg
               125                 130                 135
Gly Arg Arg Trp Ser Val Glu Leu Lys Arg Ser Val Ser His Ala
               140                 145                 150
Val Asp Val Val Leu Glu Lys Ser Glu Glu Leu Val Asp His Phe
               155                 160                 165
Leu Pro Met Thr Glu Glu Glu Leu Ala Ala Leu Ala Ala Glu Ala
               170                 175                 180
Glu Gly Pro Glu Val Gly Ser Val Glu Asp Gln Arg Arg Gln Gln
               185                 190                 195
Gly Tyr Phe Val Arg Leu Gly Ser Leu Ser Ala Arg Ile Arg His
               200                 205                 210
Leu Ala Tyr Glu His Ser Val Gly Lys Leu Arg Gln Ser Lys His
               215                 220                 225
Arg Ala Gln Asp Thr Leu Ala Gln Leu Gln Glu Thr Leu Glu Leu
               230                 235                 240
Ile Asp His Met Gln Cys Gly Val Thr Pro Thr Ala Pro Ala Arg
               245                 250                 255
Pro Gly Lys Val His Glu Leu Trp Gly Glu Trp Gly Gln Arg Pro
               260                 265                 270
Pro Glu Ser Arg Arg Arg Ser Gln Ala Glu Leu Glu Thr Leu Val
               275                 280                 285
Leu Ser Arg Ser Leu Thr Gln Glu Leu Gln Gly Thr Val Glu Ala
               290                 295                 300
Leu Glu Ser Ser Val Arg Gly Leu Pro Ala Gly Ala Gln Glu Lys
               305                 310                 315
Val Ala Glu Val Arg Arg Ser Val Asp Ala Leu Gln Thr Ala Phe
               320                 325                 330
Ala Asp Ala Arg Cys Phe Arg Asp Val Pro Ala Ala Ala Leu Ala
               335                 340                 345
Glu Gly Arg Gly Arg Val Ala His Ala His Ala Cys Val Asp Glu
               350                 355                 360
Leu Leu Glu Leu Val Val Gln Ala Val Pro Leu Pro Trp Leu Val
               365                 370                 375
Gly Pro Phe Ala Pro Ile Leu Val Glu Arg Pro Glu Pro Leu Pro
               380                 385                 390
Asp Leu Ala Asp Leu Val Asp Glu Val Ile Gly Gly Pro Asp Pro
               395                 400                 405
Arg Trp Ala His Leu Asp Trp Pro Ala Gln Gln Arg Ala Trp Glu
               410                 415                 420
Ala Glu His Arg Asp Gly Ser Gly Asn Gly Asp Gly Asp Arg Met
               425                 430                 435
Gly Val Ala Gly Asp Ile Cys Glu Gln Glu Pro Glu Thr Pro Ser
               440                 445                 450
Cys Pro Val Lys His Thr Leu Met Pro Glu Leu Asp Phe
               455                 460
```

<210> 11
<211> 3150
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7247177CB1

<400> 11
aaataaatca ttgtagaaac gctagtttgg gcctgaaaaa ttccaggagc aagagtcaag 60

atttgtcact ccatgagaat ctggagggga ctcccttccc agaaacttga cgatgaagta 120
ctggttgtaa ttttagaaag acacccaatc ggctttttta aaagatcgcc cagggccctt 180
gtcctgagag ctgggagctg gtcggagtga cagagaagcc atggaagcag cagtgaaaga 240
ggaaataagt gttgaagatg aagctgtaga taaaaacatt ttcagagact gtaacaagat 300
cgcattttac aggcgtcaga aacagtggct ttccaagaag tccacctatc gggcattatt 360
ggattcagtc acaacagatg aagacagcac caggttccaa atcatcaatg aagcaagtaa 420
ggttcctctc ctggctgaaa tttatggtat agaaggaaac attttcaggc ttaaaattaa 480
cgaagagact cctctaaaac ccagatttga agttccggat gtcctcacaa gcaagccaag 540
cactgtaagg ctgatttcat gctctgggga cacaggcagt ctgatattgg cagatggaaa 600
aggagacctg aagtgccata tcacagcaaa cccattcaag gtagacttgg tgtctgaaga 660
agaggttgtg attagcataa attccctggg ccaattatac tttgagcatc tacagattct 720
tcacaaacaa agagctgcta aagaaaatga ggaggagaca tcagtggaca cctctcagga 780
aaatcaagaa gatctgggcc tgtgggaaga gaaatttgga aaatttgtgg atatcaaagc 840
taatggccct tcttctattg gtttggattt ctccttgcat ggatttgagc atctttatgg 900
gatcccacaa catgcagaat cacaccaact taaaaatact ggtgatggag atgcttaccg 960
tctttataac ctggatgtct atggatacca aatatatgat aaaatgggca tttatggttc 1020
agtaccttat ctcctggccc acaaactggg cagaactata ggtattttct ggctgaatgc 1080
ctcggaaaca ctggtggaga tcaatacaga gcctgcagta gagtacacac tgacccagat 1140
gggcccagtt gctgctaaac aaaaggtcag atctcgcact catgtgcact ggatgtcaga 1200
gagtggcatc attgatgttt ttctgctgac aggacctaca ccttctgatg tcttcaaaca 1260
gtactcacac cttacaggca cacaagccat gccccctctt ttctctttgg gataccacca 1320
gtgccgctgg aactatgaag atgagcagga tgtaaaagca gtggatgcag ggtttgatga 1380
gcatgacatt ccttatgatg ccatgtggct ggacatagag cacactgagg gcaagaggta 1440
cttcacctgg gacaaaaaca gattcccaaa ccccaagagg atgcaagagc tgctcaggag 1500
caaaaagcgt aagcttgtgg tcatcagtga tccccacatc aagattgatc ctgactactc 1560
agtatatgtg aaggccaaag atcagggctt ctttgtgaag aatcaggaag gggaagactt 1620
tgaaggggtg tgttggccag gtctctcctc ttacctggat ttcaccaatc ccaaggtcag 1680
agagtggtat tcaagtcttt ttgctttccc tgtttatcag ggatctacgg acatcctctt 1740
cctttggaat gacatgaatg agccttctgt ctttagaggg ccagagcaaa ccatgcagaa 1800
gaatgccatt catcatggca attgggagca cagagagctc cacaacatct acggttttta 1860
tcatcaaatg gctactgcag aaggactgat aaaacgatct aaagggaagg agagaccctt 1920
tgttcttaca cgttctttct ttgctggatc acaaaagtat ggtgccgtgt ggacaggcga 1980
caacacagca gaatggagca acttgaaaat ttctatccca atgttactca ctctcagcat 2040
tactgggatc tctttttgcg gagctgacat aggcgggttc attgggaatc cagagacaga 2100
gctgctagtg cgttggtacc aggctggagc ctaccagccc ttcttccgtg gccatgccac 2160
catgaacacc aagcgacgag cgccctggct ctttggggag gaacacaccc gactcatccg 2220
agaagccatc agagagcgct atggcctcct gccatattgg tattctctgt tctaccatgc 2280
acacgtggct tcccaacctg tcatgaggcc tctgtgggta gagttccctg atgaactaaa 2340
gacttttgat atggaagatg aatacatgct ggggagtgca ttattggttc atccagtcac 2400
agaaccaaaa gccaccacag ttgatgtgtt tcttccagga tcaaatgagg tctggtatga 2460
ctataagaca tttgctcatt gggaaggagg gtgtactgta aagatcccag tagccttgga 2520
cactattcca gtgtttcagc gaggtggaag tgtgatacca ataaagacaa ctgtaggaaa 2580
atccacaggc tggatgactg aatcctccta tggactccgg gttgctctaa gcactaaggg 2640
ttcttcagtg ggtgagttat atcttgatga tggccattca ttccaatacc tccaccagaa 2700
gcaatttttg cacaggaagt tttcattctg ttccagtgtt ctgatcaata gttctgctga 2760
ccagaggggt cattatccca gcaagtgtgt ggtggagaag atcttggtct taggcttcag 2820
gaaggagcca tcttctgtga ctacccactc atctgatggt aaagatcagc ctgtggcttt 2880
tacgtattgt gccaaaacat ccatcctgag cctggagaag ctctcactca acattgccac 2940
tgactgggag ggttcctaaa aaagctctaa aatgaatgca tagtgtgtct ttccttcaag 3000
atgtaaatgt caaggaatga gttggcaaat cagtttaaaa acacaaacta cccaacagaa 3060
aaatccagaa gttccccatc cagttcgtct ggaagagacc tccggaaata tgcgcttgcc 3120
agcacactgc ggccgtaacg tgagcggccc 3150

<210> 12
<211> 3290
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5596327CB1

<400> 12

ggtctaaggt agcgagcggc tttgctgctg ctgctgcttc tggggcggcg ctgtgccgcg 60
cggcgccgcc cgggttgtct gctgctgctg ctgctggggg tctgtccgcc gggctgcggc 120
aggcgccctg gccaccgagc actactcgcc gtctccctgc tcaagcagga gctgcagacc 180
ggcagagagg aggccccggg gcgggggget gagcccgagt ccgggggact ggggggacca 240
gtactctgcc gagtgcggcg agtcatcctt tttgaacttc catgactcag actgcgaacc 300
caagggatca tcaccctgtg actccttgct ttccctcaac actgagaaga ttctgagcca 360
ggccaagtct attgcagaac agaagagatt cccgtttgcc actgataatg acagcacaaa 420
tgaagagtta gctattgctt atgtcttgat tggcagtggt ctgtatgatg aagcaatacg 480
gcatttttca acaatgcttc aggaggagcc tgatctggtt agtgcaattt atggccgagg 540
gatagcctat ggaaagaagg gactacatga cattaagaat gctgagcttg ctctgttcga 600
actgagccga gtaattacct tggaaccaga tcgtccagag gtatttgagc agcgagcaga 660
aattctgtcc cctctgggac gaattaatga agcagtgaat gacctcacta aagctatcca 720
actgcagccc tcagcacggc tgtacagaca tcggggaacc ctgtacttca tatcagagga 780
ctatgcaaca gcccatgaag actttcagca gtccttagaa ctgaacaaaa accagcctat 840
agctatgcta tacaaaggtt taactttctt tcacagagga cttctgaagg aagctattga 900
atccttcaaa gaagctttga agcagaaagt tgactttatt gatgcatata aaagtctagg 960
gcaggcatat agagaactgg gcaattttga agcagccact gagagctttc aaaaggcact 1020
gttgctcaac caaaatcatg tgcaaaccct ccagctccgg ggaatgatgc tctaccacca 1080
cggcagctta caggaagccc ttaagaactt taagcggtgt ctgcagctag agccatataa 1140
tgaagtgtgc cagtatatga aagggctcag ccatgttgcc atgggacagt tttatgaagg 1200
gataaaagca caaacaaaag ttatgctgaa tgatcctctc ccaggccaga aggctagccc 1260
tgagtatctt aaagtaaagt atctccgaga gtattctcga tatcttcatg cacaccttga 1320
taccccctt acggaatata acattgatgt ggatctgcct ggaagcttta aggaccactg 1380
ggctaaaaat ttgcctttcc tcatagaaga ctacgaagag cagccagggt tgcaacccca 1440
cataaaagat gtgttacatc agaattttga gagttataag cctgaagtac aggagctgat 1500
ttgtgtggct gatcgtttgg gatccctgat gcaatatgaa acacctggtt tcctgccaaa 1560
caagagaata cacagagcta tgggtttggc cgcattggag gtcatgcaag ccgtgcagcg 1620
tacatggacc aactcgaaag ttcgaatgaa tgggaagaca cggttgatgc agtggagaga 1680
catgtttgac attgcagtta aatggagaag gattgctgac ccagaccagc ccgtgctgtg 1740
gttagatcaa atgccagcac gaagtcttag cagaggtttt aacaaccaca ttaatttaat 1800
caggggacag gtgatcaaca tgagatacct agaatatttt gagaaaatac ttcattttat 1860
taaagacaga attcttgttt atcatggagc taataatcct aaaggattgc tggaagttcg 1920
ggaagccctg gaaaaggtac acaaagtaga agaccttctt ccgattatga agtttaatat 1980
taaaacgaag gatgggttca ccgtgaacac aaaagttccc agccttaaag accaagggaa 2040
ggaatatgat ggattcacaa tcacgattac aggagacaaa gttggcaata tattattttc 2100
tgtggaaact caaaccacgg aagaaaggac acaattatat catgctgaaa tagatgcact 2160
ttataaagat ttgacagcaa aaggaaaagt attgattctt tcatcagaat ttggggaggc 2220
tgatgctgtc tgcaacttaa tcttatcctt agtttattac ttttataatt taatgccact 2280
ctctcgagga tccagtgtaa ttgcttactc ggtcatcgtg ggagcactga tggcaagtgg 2340
aaaagaagta gcaggaaaaa ttcccaaagg gaagttagtc gactttgaag ctatgacagc 2400
ccctggttca gaggccttta gcaaagtcgc caaaagctgg atgaacttga aaagtatttc 2460
accttcttat aagactcttc catcagtttc agaaacgttt ccaacgttaa gatcgatgat 2520
tgaggtgcta aacacagact cttctccacg ttgtcttaag aaactctagt tctgctgctg 2580
tatttataca agtaaagggc cggacctctt gcttcttaag ttatttttta aaacatggaa 2640
ttataaagaa attaggtcct ctattacttt tgataccaat ttttatagga attgaatact 2700
tgaaatcatt ttctttactt ttctgaacca taatgaaaat catgaaagag gctttcgggg 2760
ggaaaaagct acttgacttg gagggaaagc gtatgtatta gccggcagcc tgttggttcc 2820
cacttataaa gtcatataac tccttcctta cagtaaaaaa aattttcctc aacaactaaa 2880
atataatact gaagatgaac aaaaacaaaa ccaaatagga gtgtttgtat tcctttagag 2940
atctggatat ttttcaaaga cttgcttttt acgcacattg tgatagagca tcttttgttg 3000
actcacgtaa cctccatgtg tgggtccttt gttttattag aaccatggaa gaaggttcct 3060
gggccatcag tatgtgttgc agagctcaca gccaggaaca cttattttga acaatgcaga 3120
cagtacttgc cagagtccca aatgctgaat tattgaacca tgtttttttt ttcctttgtt 3180
gaaaaaaata aaccaggtga tttctataaa tcaccagttt ttggaaaaaa aaaaaaaaag 3240
ggggccgcgg aaaagggagc tcgtgaaccc gggaataaat ccgggaccgg 3290

<210> 13
<211> 4279
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 4179651CB1

<400> 13
ggagaagcgg ctcctgcagg cggtggcgct ggtgctggcg gccctggtcc tcctgcccaa 60
cgtggggctt tgggcgctgt accgcgagcg gcagcccgac ggcacccctg tgtgggatcg 120
ggggcgtgtg tgtgtgctgc cgtgctggtg gtgacatggg ctcacacatg tcgacaaaat 180
gaaaacgttt ttcttgtgga gatgggcaga agctgaagtg actggcatga caatggaggc 240
catccggatg ggacgctcag cgcgtaggaa atggagaaca aggaagacct taccccatga 300
ccgatgctga gagagtggat caggcatacc gagaaaatgg atttaacatc tacgtcagtg 360
ataaaatctc cttgaatcgc tctctcccag atatccggca cccaaactgc aacagcaagc 420
gctacctgga gacacttccc aacacaagca tcatcatccc cttccacaac gagggctggt 480
cctccctcct ccgcaccgtc cacagtgtgc tcaatcgctc gcctccagag ctggtcgccg 540
agattgtact ggtcgacgac ttcagtgatc gagagcacct gaagaagcct cttgaagact 600
acatggccct tttccccagt gtgaggattc ttcgaaccaa gaaacgggaa gggctgataa 660
ggacccgaat gctgggggcc tcagtggcaa ctggggatgt catcacattc ttggattcac 720
actgtgaagc caatgtcaac tggcttcccc ccttgcttga ccgcattgct cggaaccgca 780
agaccattgt gtgcccgatg attgatgtaa ttgaccatga cgactttcgg tacgagacac 840
aggcagggga tgccatgcgg ggagcctttg actgggagat gtactacaag cggatcccga 900
tccctccaga actgcagaaa gctgacccca gcgacccatt tgagtctccc gtgatggccg 960
gtggactgtt cgccgtggat cggaagtggt tctgggaact cggcgggtat gacccaggct 1020
tggagatctg gggaggggag cagtatgaaa tctccttcaa ggtgtggatg tgtgggggcc 1080
gcatggagga catcccctgc tccagggtgg gccatatcta caggaagtat gtgccctaca 1140
aggtcccggc cggagtcagc ctggcccgga accttaagcg ggtggccgaa gtgtggatgg 1200
atgagtacgc agagtacatt taccagcgcc ggcctgaata ccgccacctc tccgctgggg 1260
atgtcgcagt ccagaaaaag ctccgcagct cccttaactg caagagtttc aagtggttta 1320
tgacgaagat agcctgggac ctgcccaaat tctacccacc cgtggagccc ccggctgcag 1380
cttgggggga gatccgaaat gtgggcacag ggctgtgtgc agacacaaag cacggggcct 1440
tgggctcccc actaaggcta gagggctgcg tccgaggccg tggggaggct gcctggaaca 1500
acatgcaggt attcaccttc acctggagag aggacatccg gcctggagac ccccagcaca 1560
ccaagaagtt ctgctttgat gccatttccc acaccagccc tgtcacgctg tacgactgcc 1620
acagcatgaa gggcaaccag ctgtggaaat accgcaaaga caagaccctg taccaccctg 1680
tcagtggcag ctgcatggac tgcagtgaaa gtgaccatag gatcttcatg aacacctgca 1740
acccatcctc tctcacccag cagtggctgt ttgaacacac caactcaaca gtcttggaaa 1800
aattcaatag gaactgagcc ctcatgtccc cttggcaggc cccccagggt ctggcactca 1860
ctgcagactt cctctttcaa gggaggcagg gcccctgtgg gcactaggtg taaaaggtgc 1920
tggccaaatg gttcagggtg aagagggctc ttgattcagg ggctggggtc tgcctggtcc 1980
ttgagcccct gagttgtggg ggtagggtga agagcatatc ccacaagagg ccccacaggg 2040
agcagagact gctttaatcc ctgctgacat cacggaaaag caacagagcc ttttcaactt 2100
tgtcactatg tccccttgaa cattatgtgg gagaacacca aggtagccta ggccacccaa 2160
aagtgagtcc tgcgaggttg cccagccctc agatggctct cctacatgat ggtgctttag 2220
aaacaaaggt aaaatttgcc tgtttggggc agcttttagt atcgatgcca ctcatctgca 2280
gcagaagaga aagaagtcct cttggggctt tttagtttct gccgtcctgg ggggaacatt 2340
gcagttactg cacagcttct gttctctgtc acaaccccag gtgatttggt ccggtcaaag 2400
gccatacttg gggccctaag agtgttcagt attgaatgct gatcagctgc caggtgagga 2460
gtcagaagag ggagcccccc tagacatttc tttgcagcta tggacatgcg ggatatctcc 2520
ccctgctctc tgggtatttg aaatgtcaat tttagcactc tccaggcaca aggacagccc 2580
agcaccagct ttacagggca gtgtttcaga tggccctgag cccacggaaa aggccaggta 2640
gacctccaaa ctagaaatgc tggctgattt gccctgatcc atgcttccat ttccctgtct 2700
ctcttcccca ggcaattact ggcctcaaaa gaggaacaga ggtgctgcga ggtgctcacc 2760
tcacagagtc tggaggcctc caggatcaac tgtgggcaaa gtgcctgcct ctgacctcat 2820
catggttcta gttctcatac agaactccag aatttttaaa gaactctata attggattgc 2880
aaactaggat gctacatagg attctggtat tccacatcca atatggattt ctagaatgct 2940
gtgattaaag gagccagcca ggtgtaatac agtcaaggca gcccccagcc tagagacaat 3000
ctgtgaaatc caaagttggt ggtgttggga aagcaggggg acatgtgtcc ctcagctcag 3060
cagaggctgt ggtacaacat ggtccttggt gaagacctgc acccctggaa cctcccacca 3120
tcatcacaac tgtagtctca tttgcagtgg agaaaagaac ccgacgtccc acagccagat 3180
atacacccag ctccatgcca gcccttcatg tttacctttt gctttgttaa ttacatgtca 3240
gactcctaga gggcctccag actaatagga agcatttctg taaccaacct gccacccact 3300
gattcagaaa tggaaatcac attccacaat ctatggcttc caccagctag cccaggaaat 3360
acttgaaatc agcattccaa ttagtgttga gtctcttgat tgtgtcattt accaattaaa 3420
taactgagac ctaagtctgg gaacagagcc acgaatctgc ctttgagatg ctggcagatc 3480
tcaaggccat caattattgg gggagggagg gacaaacact cccaatcatc caccagtcag 3540
actgaatgtg tagctggcga ggaattactt ccacttctgg cccagcacaa gccctgcttt 3600

ggccacctgt ctgcaagaga ggcggcccct gtgcttgcaa cgcttacgtg ttgatcccag 3660
tgtccttttc caaatgagtg ctgtagcttt agaagtggcc ctctatagaa agaagtcaaa 3720
agatgaggcc ccttctagaa tctaggataa caagagtgtt gacagtttga ggagtcgaat 3780
tgagattcat catcaaagag caatgcagcg tcgttaaaat aaaaactgtg ccttttaaaa 3840
agaaaaatgc aaatatagag caaatcccta aacttgaacc atttcctagt gccttgctag 3900
acaaacataa aggggcaggg ttgtggggag gggaacgttt ttggtggtgt gtgcagttcc 3960
tactggatga gtgtgtgttt cttaatgtct catttcaagc aggagatgtt gggtctggag 4020
caagatctga gactgagatg tcccccaagg gtgacaggtc agatttattt cagtcagtgc 4080
aagtcacact ccaaactaaa ggctgggcag tgcgtttagt ctgtggccta gaagacctca 4140
tcagcccccg agaggaggcc ttgcattacc tcactgggac ctgtttgggg gccatccctg 4200
cggtccaccc tctgaacccc cagcatgttt gtcctgtttc tcaccattgg gttaaggggt 4260
gccgagcact agctagact 4279

<210> 14
<211> 1042
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7482109CB1

<400> 14
atgaattcta aaagaatgct gttattggtt ttatttgctt tttcactgat gttggttgag 60
cgttatttca ggaatcacca agtagaagaa cttcggctct cagactggtt tcatcctaga 120
aaacgccctg atgttataac gaaaacagac tggctcgctc ctgtcctatg ggaagggact 180
ttcgacaggc gggtcctgga aaaacattac agaaggcgga atatcactgt gggcctggcc 240
gtctttgcta ctggcaggtt tgcagaggag tacctgaggc cgttcctaca ctccgcaaat 300
aagcacttca tgacaggcta ccgagtgatc ttctacatca tggtggacgc cttcttcaag 360
ctgcctgaca tagagcccag tcctcttcga acgttcaaag catttaaagt gggcaccgag 420
aggtggtggc tcgatggccc cctggtgcat gtgaagagcc tgggtgaaca catcgccagt 480
cacatccagg acgaggtgga cttcctcttc agcatggctg ccaaccaggt cttccagaat 540
gagttcgggg tggagaccct gggcccgttg gtggcccagc tccacgcctg gtggtatttc 600
agaaacacca agaacttccc ttatgagagg aggccgacct cagcagcttg catcccgttt 660
ggacagggag atttctatta tggcaacttg atggttggtg gcacacccca taatatttta 720
gacttcatca aagaatatct gaacggagtt attcatgaca tcaaaaatgg actcaatagc 780
acttatgaaa agcaccttaa caaatatttt tacctcaata aacccaccta gctgttatca 840
ccagcataca gctgggatct tgcattttct cctcctccac agatccaata cgtcaaggtc 900
gcacatgatt cccagaggaa attatgaatt acgccatgag gtttatgacg aataaatgaa 960
tcacggcaat ttcttataaa tgagggaagt cttaagacct tccatttttt agagacatac 1020
aaatttacat cctccccaaa aa 1042

<210> 15
<211> 4320
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2867618CB1

<400> 15
attcatttgg ggaaaaattt tgatttccaa agagtgacag aatcaacagt gaaacgaata 60
ccaagaattt aaaaagtgtt gaaatcacta tgaaaccttc caaggcctct gaacttaact 120
tggatgaact accacctctg aacatttatc tacatgtatt ttattacagt tggtatggaa 180
atccacaatt tgatggtaaa tatatacatt ggaatcatcc agtgttagag cattgggacc 240
ctagaatagc caagaattat ccacaaggga gacacaaccc tccagatgac attggctcca 300
gcttttatcc tgaattggga agttacagtt ctcgggatcc ttctgtcata gaaactcaca 360
tgagacaaat gcgctcagct tcaattggtg tactagccct ctcttggtac ccacctgatg 420
taaatgatga aaatggagaa cctactgata acttggtacc cactattttg gataaagctc 480
ataaatataa cctaaaggtt acttttcaca tagaaccata tagcatcgag atgatcaaac 540
atgtacaaaa tgtcaagtat attatagaca aatatggaaa tcatccggcc ttttacaggt 600
acaagacgaa gactggcaat gctcttccta tgttttatgt ctatgattcc tatgtgacca 660

agcctgaaaa aagggccaat ctgttaacca cctcagggtc tcggagtatt cgcaattctc 720
cttatggtgg actgtttatt gcccttctgg tagaagaaga acataagtat gatattcttc 780
aaagggttga tggaatttac acatattttg ccacaaatgg ctttacttat ggctcatcac 840
atcagaatgg ggctagccta aaattattgt gtgataaata caacttaata tttatcccaa 900
gtgtgggccc aggatacata gataccagca tccgtccatg gaacacgcaa aacactcgga 960
accgaatcaa tgggaagtat tatgaaattg gtctgagtgc cgcacttcag acacgcccca 1020
gcttaatttc tatcacctct tttaatgagt ggcatgaagg aactcagatt gaaaaagctg 1080
ttcccaaaag aaccagtaat acagtgtacc tagattaccg tcctcataaa ccaggtcttt 1140
acctagaact gactcgcaag tggtctgaaa aatacagtaa ggaaagagca acttatgcat 1200
tagatcgcca gctgcctgtt tcttaatgca ttgattaaag tttaatagtt atcaaaatca 1260
cctaatttt aaaaatagct ttcgttttga gttctggaaa gaaaactgtc aaaatcagta 1320
tatactatta gttatattta aaaatatttt tttaaattct ttacagataa tattatactt 1380
gttacccttc acaataccac atgagaaaat atctgagaca aaatgtatac aaatatattc 1440
cttatggcat aatttattgc atttctgact gaaatcaaaa ttctgatttg atggcaattg 1500
aattttcatt ttacaataga taaatgcttg tgttacctaa agcacttagc acacagttaa 1560
attatattta catcctagac ccaaataaat aggattgtgt gtatatttgg gatatctatt 1620
gaagaaaaaa agaaaacccc ttaaagataa tgtacatgct tcatgtcatg tctttaaaat 1680
aatttaatca actttattgt cttagtattt agactctgga taactctaca ataatgagga 1740
aattcttaag aataacaaaa tcactgtacc ttcctctcaa ttttgctgtg aacctgaaat 1800
ggctttaaat taatactctt atttttatt taatttaatt acataaatta aaccttacca 1860
tgaccaaatt gtgttaggaa ggcctgctat ctacagcaca gtgtgtcatt tgcagatttg 1920
tggttaccta taccacgcta ggtgttttga catgtttagt gtttctgctt tacagtgctg 1980
aattccatat tttagaagct atgaaagtcc ttttatgaaa aagttactga ttgcttctca 2040
gttattagga aaacagttgt ttcacaatta ttatgtagat atgatgccca aatatcattt 2100
ttagtatatc ttgtcgatct ttaagttgtt actattgtgt tattcatgtc tttaaatcag 2160
ataccaaata ttttttagga aagaaaaatg ttattactgt cattaggttg tcttttaata 2220
ctttaagtta ttttgacgaa aagtaataga gaaaatttac ttagcatttt agattctaga 2280
gacatggaaa tgaaaattat tttatgtcta gagtaggtcc tgaagtttgg ctttacatta 2340
agtttagcac tgtatcagaa tgaagaaact aatattttac ataaaaacta atactttcaa 2400
ttttttatat agtaatatcc ccattttgta aatgttagac ttttatcata cctgtaagtt 2460
aaaatacttg ttatcaataa cttgtcatgt tgtgacaaat tgatcacttg tgtacgaaaa 2520
ataaatctcc ttaaaaacta aataaaatgc actgtattct tacagttaat gtttataact 2580
atagtaaaaa attaatatat atcctattac ataaatgtta tttcttaggt gttccattaa 2640
gaagagcaat agaataatgc taaaaaataa tgcctataaa tcttcagagt ataaagacat 2700
ccattcagaa acaaaaatta gcactaaatt ttttataaaa tagaccagat gacaaaattt 2760
attttatttt taaacagtgg ttttgacaca aattatgtta ttgaaaagca ttattaatgt 2820
ttaatttatt taaaattttg gaatttgcca tttctcagac aatgatcagg ccttaggaaa 2880
ttaatacagt agtagtaatc attttctagg ggaaaataaa agaataaatc actatactga 2940
tattttgata taagcaagca cttacatggt aatcactata tagatccaac ctgtggattt 3000
tcttcttatg tccatttaac tagaatatat tattttaggt ataatttaca aatgtcacac 3060
ctaataatct tttataatat accatatttc attaaagttt tgttagagaa gtatctacca 3120
cagaggagtt tttgtcattg tgtacgttgt gtatttgaac ccaccatgac agaaagtaaa 3180
ttttaggaaa tagttatgag attaagggaa aatctataaa aacaaggtta gcatattctc 3240
aacacagata ccaccacttt ctttttccca ttatagacat ggtgaatcca cacagcatac 3300
ttcatctctg agctttgttg tgattcctca acacattacc ctaaccagcc agcagtaaca 3360
gatttcagag taagataaag cagattctgt cttcattgca aaaagttatt ctcaatggaa 3420
gaatggcatc tgatctcata attactagtt tatattaata tagtttttt ctccctttt 3480
aataaaataa ttacagtcat ccctcagtgt ctgtggggga ttggttccag ttacccctat 3540
agataccaaa atctgcagat gctcaagtcc ctgatataaa ctggcatagt agttgcatat 3600
aatctatgca catcctcctg tatacattaa gtcatctcta gattatttat aacacttaat 3660
acaatgtaaa tgctatgtag ttgttatacc atactggtta gtgaataatt acatgaaaaa 3720
aaagagtctg tacatcttca gagtttcagt cggcaatttc ttggccatgg atgtagaacc 3780
tacagataag gtgagccaac tgcattagga aataactcta ataattctgt taattcttag 3840
agaggaaaac tttcaaaatc ttcctcaggt atttattaca actgccttta ccattttagt 3900
tgtaacacag tttaaattgt tatgataaca agtaaataag agcaaagaat ttatttctta 3960
attcaaaact atacgtttga attcaatatg gtataactta aagtggtata atacatacaa 4020
tgcatgaatc ataatggatt cttttataag ttattaattt ttatggttta atcagtctaa 4080
ttgttttgac tgttatagaa accaaatatt ttactgtttc ttttaaggac taatattgtc 4140
aaaaactgct gttattaact tcacttgagt tgtttaactt ccttctgttt taagattgta 4200
attaaaaatt actattttgt tatatggaat ggttaatttt tacctaataa aaacatagat 4260
gaaatacatt gtattttaga tattaaagtc tatcatttgc tctttaaaaa aaacaaaaaa 4320

<210> 16

<211> 1733
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7488348CB1

<400> 16
atcggatctg agcgaacgga acgggtgcgg gtgttcaaga tgtccattct caagatcatc 60
catgcccgtg atatctttga atcccgtggg aatccaactg ttgaggtaga cctttacact 120
aacaagggtg gtctgttcgg aagagctgct gttcctagcg gtgcctcaac tggaatttat 180
gaagctctgc tggaacttcg ggacaatgac aagacacgct acatgggggg gaagggtgtc 240
tcaaaagctg ttgagcacat tatcaataaa acaattgcac ccgcactgat tagcaagaat 300
gtcaatgtag tggagcaaga caagatcgat aacctgatgc tggacatgga tggatcagag 360
aacaaatcta aatttggtgc caacgccatc ttgggtgtat ctctggctgt atgctcgaat 420
gctggtgcta ctgctgagaa gggtgtcccc ttgtaccgtc acattgctga ccttgctggg 480
aataacccag aagtcatcct gcctgttccc gcttttaacg tgatcaacgg tggctcccat 540
gctggcaata agctggctat gcaggagttc atgatccctc cctgtggtgc tgacaggttt 600
aacgacgcaa tccgcattgg tgcagaggtt taccacaacc tgaagaatgt catcaaggag 660
aaatatggga aagatgccac caatgtgggg gatgaaggcg ggtttgctcc caacatcctg 720
gagaataaag aaggcctgga gctgctgaag actgctattg ggaaagctgg ctacactgat 780
aaggtggtca tcggcatgga cgtagcggcc tccgagttct tcaggtctgg gaagtatgac 840
ctggacttca agtctcccga tgaccccagc aggtacatct cgcctgacca gctggctgac 900
ctgtacaagg gctttgtcct ggggcatgca gtgaagaact acccagtggg cgtctccatc 960
gaggacccct catttgacca ggatgactgg ggtgcctgga agaagctgtt tactggcagc 1020
ctggttggca tccaagtggt tggagacgat ctgaccgtga ctaagcccga agcgcgtatt 1080
gccaaggctg tggaggaggt taaagcctgc aactgcctcc tcctcctcaa ggtcaaccag 1140
attggatctg tgacggagtc cctacaagct tgcaagcttg cccagtccaa cggctggggc 1200
gtgatgcccg tgagtcaccg cctctccgga gaaacagaag ataccttcat ggctgacctc 1260
gtggtcgggc tctgcactgg tcagatcaaa actggtccca cttgccgatc tgagcgtcta 1320
gccaagtaca accagctgct gaggattgaa gaggctgagg ctggcagcaa ggcccgcttt 1380
gctggaagaa acttcaggaa cccccgtatc aactaagctg cgtggatcga cacccgttct 1440
ggttatgtaa gcactagtca cctacttaga ctcacaatta cttgtattag aagatgaggg 1500
gcaggctgaa gaaaagacca gtttgcaggt cctctcccct cctagatgac tctccttcac 1560
ctacagtgtt tccaccagct ctgatctgtt acttgtaacg atcgtgcttt gtagaacaat 1620
cccagtcttt gatgtttggg aggctgtttg acttgcagga cagcaagacg gtacctacaa 1680
aacagctagt agtgttttta catgtgataa ataaaaagca tcaaacaaaa aaa 1733

<210> 17
<211> 2201
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5539166CB1

<400> 17
gacagggtta agcaattttt ttaaagcact acatatttat tgatagaata tctctccagg 60
cagtattcac atgctttaaa gtgaagctcg tagcttggca gtgcagttta agcaaataaa 120
ctcaggctgt tgaacacaat gattaacagg aagcagggga ggggacatag ttactagact 180
ctggccggga caagaggctg aggttaaggt cttagagccg atgacttgca gaatggtcac 240
ctcacccact tgcaatttca tggtgggaac aggtggaccc cctggaatca gaacctctct 300
gaggacatct gtttttgtgt agacacaggt tgcaggttag caggagaaca ggcaagccaa 360
atgcaaagga gccacttcag aaatgtgtca cagaaaagtg aaaatgcaac ctagtggtaa 420
gtgaagaggg gaagaagaaa gaaaaaggac cagaaccgtg aactgaaggg acggggaaca 480
gccagacgag agcttcagcc atcacgagga tgatttcgga acctggagaa aatgtaagtt 540
aaatatatct acactctgat cctatctcaa gagagagata ttttactcat ttcctggttg 600
tgaatgatgg gctcttggaa gcactgtctt tttagcgcgt ctcttatctc tgccctgatt 660
tttgtatttg tttacaatac tgagttatgg gagaataaac gttttctgag ggcagctctg 720
tccaatgctt cactgttagc agaagcctgt catcagattt ttgaggggaa agttttttac 780
ccaacagaaa atgcattgaa aactaccctt gatgaagcta cctgctatga gtacatggtt 840

cgaagccact atgtaacaga aacactctct gaagaagagg ctgggttccc tttagcttac 900
acagtgacca tccacaaaga cttcggcact tttgagaggc tcttcagggc gatttatatg 960
ccccaaaatg tctactgtgt gcacctggat cagaaggcga cggatgcctt taaaggtgca 1020
gtgaaacagt tactcagctg cttcccaaat gcttttctgg cttccaagaa ggagtcggtt 1080
gtctatgggg ggatctccag gctccaggct gacctgaact gcctggaaga ccttgtggcc 1140
tctgaagttc cctggaagta tgtcatcaac acctgcgggc aagactttcc cctgaaaacc 1200
aacagggaaa tagttcagta tctgaaggga tttaaaggga aaaatatcac ccccggagtg 1260
ctgcctcctg accacgctgt tggacggact aaatacgtcc accaagaact gttaaaccac 1320
aaaaattcct acgtgattaa aacaacaaaa ttaaaaactc ctcctcctca tgacatggtg 1380
atttaccttg gcacggccta cgtggctctc acaagggact ttgctaactt cgtcctccaa 1440
gaccagctcg cacttgactt actctcctgg tccaaggaca cctacagccc cgacgaacat 1500
ttctgggtga cactcaacag gattcccggt gttcctggct ctatgccaaa tgcatcctgg 1560
actggaaacc tcagagctat aaagtggagt gacatggaag acagacacgg aggctgccac 1620
ggccactatg tacatggtat ttgtatctat ggaaacggag acttaaagtg gctggttaat 1680
tcaccaagcc tgtttgctaa caagtttgag cttaatacct acccccttac tgtggaatgc 1740
ctagaactga ggcatcgcga aagaaccctc aatcagagtg aaactgcgat acaacccagc 1800
tggtattttt gagctattca tgagctactc atgactgaag ggaaactgca gctgggaaga 1860
ggagcctgtt tttgtgagag acttttgcct tcgtaatgtt aaccgtttca ggaccacgtt 1920
tatagcttca ggacctggct acgtaattat acttaaaata tccactggac actgtgaaat 1980
acactaacag gatggctggg tagagcaatc tgggcacttt ggccaatttt agtcttgctg 2040
tttcttgatg ctcacctcta tattagttta ttgttaggat catgataatt taatgacctc 2100
agatcttggg cgatactctt cttccgggtt ttctcaccgg gatgtgttct ctgctcgcat 2160
acaatcgtcc tccgagggct tttaccttca gcccggggcc c 2201

<210> 18
<211> 457
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500341CB1

<400> 18
gcttgcggct cgggtggctg agcgcgcggg gaaatggcca cggggacaga ccaggtggtg 60
ggactcggcc tcgtcgccgt tagcctgatc atcttcacct actacaccgc ctgggtgatt 120
ctcttgccat tcatcgacag tcagcatgtc atccacaagt atttcctgcc ccgagcctat 180
gctgtcgcca tcccactggc tgcaggcctc ctgctgctcc tgtttgtggg cagctttcct 240
taccctgaca cagccccaga ccccacaaag ccttctggac ctggaaagcc tggggaagga 300
ctgacagacc ccaggaccag ccctggggct cagggcagcc accccgggcc gctgaccgac 360
tgacctctcc tcacggaggc ccagccccaa agccccaggg ctggcccgtt tgggacagct 420
gaccaataaa cactgatggt gtgtttaaaa aaaaaaa 457

<210> 19
<211> 1513
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500846CB1

<400> 19
ggagctccga gtgtccacag gaagggaact atcagctcct ggcatctgta aggatgctgt 60
ccatgctgag gacaatgacc agactctgct tcctgttatt cttctctgtg gccaccagtg 120
ggtgcagtgc agcagcagcc tcttctcttg agatgctctc gagggaattc gaaacctgtg 180
ccttctcctt ttcttccctg cctagaagct gcaaagaaat caaggaacgc tgccatagtg 240
caggtgatgg cctgtatttt ctccgcacca agaatggtgt tgtctaccag accttctgtg 300
acatgacttc tgggggtggc ggctggaccc tggtggccag cgtgcacgag aatgacatgc 360
gtgggaagtg cacggtgggt gatcgctggt ccagtcagca gggcaacaaa gcagactacc 420
cagaggggga tggcaactgg gccaactaca acacctttgg atctgcagag gcggccacga 480
gcgatgacta caagaaccct ggctactacg acatccaggc caaggacctg ggcatctggc 540
atgtgcccaa caagtccccc atgcagcatt ggagaaacag cgccctgctg aggtaccgca 600

ccaacactgg cttcctccag agactgggac ataatctgtt tggcatctac caggggaatt 660
tgttgcagga ttcgttcagt tccgggtgtt taataacgag agagcagcca acgcctttg 720
tgctgggata aaagttactg gctgtaacac tgagcatcac tgcatcggtg gaggagggtt 780
cttcccacag ggcaaacccc gtcagtgtgg ggacttctcc gcctttgact gggatggata 840
tggaactcac gttaagagca gctgcagtcg ggagataacg gaggcggctg tactcttgtt 900
ctatagatga gacagagctc tgcggtgtca gggcgagaac ccatcttcca accccggcta 960
tttggagacg gaaaaactgg aattctaaca aggaggagag gagactaaat cacatcaatt 1020
tgcaaaaaaa aaaaaagggg gggccgctct aaagaacccc ccaaggggcc caaccttacc 1080
cttgcatgca acttcatacc tctcccccta tgtaaaaagt gagaaggggc cggggccaga 1140
acctctgggg tgcccccagc ccccggtggg tgggccgatt gaccggttct ggaaggggta 1200
accagtgggc ctttgtaaaa aggcctggag ttgcccaagg tgtgctgatg gggtaggaaa 1260
agacagtgag gaggatatac ttcagagtgc ctgagaagcg cctataaaac atccctacga 1320
agtggaaaag ccgggtgtgt gaataacacc cctgcccagc tccagaagct gcatctgtga 1380
ccgcctctaa aattcgcatg gaccgattgg aaccctcccc aaggggaagt tgttgttagt 1440
aggtgcaaca gcccagggga atctgagatc attcaccaac tgtatatccc actacctttg 1500
ttgcgatgta ctc 1513

<210> 20
<211> 2084
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3230921CB1

<400> 20
ggcgggaccg gcgggggtgt ggagactcga gcctggggtc ggcggagaca gctggtgtct 60
gaagccgctc gcgcccaggg tgaccctgtt tgcagcacga tgtctgaaga agaggcggct 120
cagatcccca gatccagtgt gtgggagcag gaccagcaga acgtggtgca gcgtgtggtg 180
gctctgcccc tggtcagggc cacgtgcacc gcggtctgcg atgtttacag tgcagccaag 240
gacaggcacc cgctgctggg ctccgcctgc cgcctggctg agaactgcgt gtgcggcctg 300
accacccgtg ccctggacca cgcccagccg ctgctcgagc acctgcagcc ccagctggcc 360
actatgaaca gcctcgcctg caggggcctg gacaagctgg aagagaagct tccctttctc 420
cagcaacctt cggagacggt ggtgacctca gccaaggacg tggtggccag cagtgtcacg 480
ggtgtggtgg acctggcccg gaggggccgg cgctggagcg tggagctgaa gcgctccgtg 540
agccatgctg tggatgttgt actggaaaaa tcagaggagc tggtggatca cttcctgccc 600
atgacggagg aagagctcgc ggcactggcg gctgaggctg aaggccctga agtgggttcg 660
gtggaggatc agaggagaca gcagggctac tttgtgcgcc tcggctccct gtcagcacgg 720
atccgccacc tggcctacga gcactctgtg gggaaactga ggcagagcaa acaccgtgcc 780
caggacaccc tggcccagct gcaggagacg ctggagctga tagaccacat gcagtgtggg 840
gtgaccccca ccgccccggc ccgccctggg aaggtgcacg agctgtgggg ggaatggggc 900
cagcgccctc cggagagccg ccgccggagc caggcagagc tggagacgct ggtgctgtcc 960
cgcagcctga cccaggagct gcagggcacg gtagaggctc tggagtccag cgtgcggggc 1020
ctgcccgccg gcgcccagga gaaggtggct gaggtgcggc gcagtgtgga tgccctgcag 1080
accgccttcg ctgatgcccg ctgcttcagg gacgtgccag cggccgcgct ggccgagggc 1140
cggggtcgcg tggcccacgc gcacgcctgc gtggacgagc tgctggagct ggtggtgcag 1200
gccgtgccgc tgccctggct ggtgggaccc ttcgcgccca tccttgtgga gcgacccgag 1260
cccctgcccg acctggcgga cctggtggac gaggtcatcg ggggccctga cccccgctgg 1320
gcgcacctgg actggccggc ccagcagaga gcctgggagg cagagcacag ggacgggagt 1380
gggaatgggg atggggacag gatgggtgtt gccggggaca tctgcgagca ggaacccgag 1440
acccccagct gcccggtcaa gcacaccctg atgcccgagc tggacttctg acccatgggc 1500
cagtggaggc ggggaggaaa ggccacctgc acaccccgat ccctgctgcc ccctggtggc 1560
cacacgtaag ctcgaggcct tggccttgac ccttctttgg aatcaggccc aactccggat 1620
ctctgaccac ctttttggta ttggactctc ccatttttc cttgaacaca tggacaaaga 1680
ggcccggggg agcagggcct cgaaccctat tcaggccaac ttgagccaca agctgggttc 1740
ttcacctatg tcctgctccc tggctccatg aagcgaatcc aaatctttcc aagaggctgg 1800
gcacagtggc tcacgcctgt aatcccagca ctttgggagt ctgaggcagg tggatcatct 1860
gaagtcagga gttcgggatc atcctggcca acatgatgaa accctgtctc tacttagaaa 1920
gacagacaaa aaaaaaaagg gggggccgc cgattaggga gctcgtggac ccggggatta 1980
aatccgggac cggcccggcg ggggtaccag ttcccctaat gggagcgtat aaagctgggg 2040
aaccatggca tacgggtcgt gtggaattgt tccgtcaatc caaa 2084